TOWARDS AN ENZYMATIC ROUTE TO 2,5-FURANDICARBOXYLIC ACID

A Dissertation
Presented to
The Academic Faculty

by

Harrison Bellow Rose

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
May 2018

TOWARDS AN ENZYMATIC ROUTE TO 2,5-FURANDICARBOXYLIC ACID

Approved by:

Dr. Andreas S. Bommarius, Advisor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. M.G. Finn
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Christopher W. Jones
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Gabriele Sadowski
Department of Biochemical and Chemical Engineering
TU Dortmund

Dr. Matthew J. Realff
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Date Approved: March 23, 2018

“Look it up…”
and
“Figure it out…”

This thesis is dedicated to my mother and father
who encouraged me to seek my own answers.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge support from the National Science Foundation (NSF) under grant No. 1336386 and under contracts IIIP-1540017 and IIIP-0969003 to the I/UCRC Center for Pharmaceutical Development (CPD). Some parts of this work (CHAPTER 4) were further supported by the German Science Foundation (DFG) grant Nos. HE7165/5-1 and SA700/20-1.

I would like to thank Professors Andreas Bommarius, Christopher Jones, Yoshiaki Kawajiri, Matthew Realff, David Sholl, and Krista Walton. Thank you for bringing me into such an interesting research project and for entrusting me with carrying out the biocatalytic component of your brainchild. Thank you also, for sticking with me patiently and supporting me even as it became apparent that the enzyme development would be more involved than anticipated. Thank you also, Professors M.G. Finn and Gabriele Sadowski for agreeing to serve on my committee and for your advisement and suggestions on the work contained herein. In particular, thank you Professor Sadowski for entering into an international collaboration with me, sight unseen. To Thorsten Greinert and Dr. Christoph Held, thank you for joining in that leap of faith. I learned a lot from our partnership, and I know that you both will continue to produce amazing work.

I would also like to thank Professors Mark Styczynski and Corey Wilson, and Dr. Jamie Davey for the many insightful and enlightening discussions that helped to shape this work.

Thank you to those who worked with me and assisted me in the lab. Nayra Sofia Romero Cuellar, the HPLC methods we developed and debugged together were used in this work and appear in CHAPTER 3 and CHAPTER 4. Madison Wilber, your impeccable work on the pH and solubility projects contributed greatly to the datasets in CHAPTER 4 and CHAPTER 5. Aimee Moise, the lessons we learned together are scattered throughout CHAPTER 3 and your efforts in enzymology will help to support and guide future iterations of this project and others like it.

Thank you, Dr. Katherine Vest (Cobine Laboratory, Department of Biological Sciences, Auburn University, Auburn Alabama), for helping with enzyme characterization. The conclusions drawn from your metals analysis are discussed in CHAPTER 3. Thank you Dr. Nils Persson, not just for your advice on the Monte Carlo simulations that appear in CHAPTER 5, but also for your friendship. I would also like to thank you, professor Christoph J. Fahrni (School of Chemistry & Biochemistry, Georgia Institute of Technology), for use of your titration apparatus and pH probe, and thanks to Dr. M. Thomas Morgan, and Dr. Adam McCallum for your assistance therein. Thank you also, Dr. Morgan, for your insights and discussions relating to thermodynamic activity and ionic strength. These contributions helped to support the work presented in CHAPTER 5.

Thank you to Professor Kurt Faber and Dr. Silvia Glueck (Austrian Centre of Industrial Biotechnology, the Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Graz, Austria), and to Dr. Karl Payne and Professor David Leys (Centre for Synthetic Biology of Fine and Specialty Chemicals, Manchester Institute of Biotechnology, The University of Manchester, Manchester, UK) for your fine

contributions to biocatalysis and for providing some of the expression constructs used in CHAPTER 3 of this work.

I would like to thank Dr. Satin Gungah (Control and Power Group, Department of Electrical and Electronic Engineering, Imperial College, London, UK) for identifying that citric acid equilibrium could be posed as an eigenvalue problem. This work appears in CHAPTER 5, and may serve as an enduring reminder that fruitful collaborations can be found in unexpected places.

Thank you to the Flavinistas – for your openness and compassion in welcoming me into the protein fold. The Atlanta Flavin Group and the Southeast Enzyme Conference have provided wonderful suggestions and feedback on the enzymology in CHAPTER 3, not to mention the sheer entertainment value of our monthly meetings.

Thank you to the past and present members of the Bommarius Research Group. Drs. Lizzette Gomez Ramos, Matt Mistilis, Yuzhi Kang, and Jungmin Oh, thank you for welcoming me into the lab family. Drs Sam Au Gee, Ryan Clairmont, Jonathan Park, and Michael Rood, thank you for the assistance both in and out of the lab and for being so generous with your time and open with your advice and suggestions.

Mariétou Paye and Aditi Sharma, thank you for opening your projects to collaboration. I learned so much from both of you, and from our collective struggle to push through the seemingly endless challenges presented not only by our research, but by grad life in general.

Bettina Bommarius, thank you for helping me grow into a comfortable role in the lab, and for your endless advice and support in debugging molecular biology.

I would like to thank you, Mick Robbins, for so many things. In the lab you taught me not just tools and techniques, but also what it really means to be a scientist and to think critically and logically about data. Outside of the lab, you have been a mentor, a confidant, and a friend.

To Adam Caparco, Robbie Franklin, Nick Kruyer, Thomas Kwok, Matt McDonald, Mike Stellato, and Shan Tie, it has been a pleasure and a privilege to learn and work at your side. Your feedback, guidance, and friendship were invaluable assets over the last few years, and I have learned so much from each of you.

Finally, I would like to thank my friends and family. These last few years have been trying, and I would not be here if not for your continued encouragement and support. Thank you for joining me on this journey. I know that we have many more adventures to look forward to on the road ahead. I love you all so very much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

General Symbols Introduced in CHAPTER 1:

$\overset{i}{\rightarrow}$	Numbered edges in a reaction network

Mathematical Symbols Used in CHAPTER 2:

$\Delta_f G^0$	Standard free energy of formation
$\Delta_r G^0$	Standard free energy of a reaction
ν	Stoichiometric coefficient
R	Gas constant
T	Temperature
K_{eq}	Equilibrium constant of a reaction

Mathematical Symbols Used in CHAPTER 4:

K_{eq}^{th}	Thermodynamic equilibrium constant
a	Thermodynamic activity
ν	Stoichiometric coefficient
C	Molar concentration
γ, γ^L	Thermodynamic activity coefficient
ΔH_{fus}	Enthalpy of melting a pure substance
T_{fus}	Melting temperature of a pure substance
$A^{residual}$	Residual Helmholtz energy
$A^{hard\ chain}$	Residual Helmholtz energy contribution due to hard-chain repulsion
$A^{dispersion}$	Residual Helmholtz energy contribution due to van-der-Waals dispersion
$A^{association}$	Residual Helmholtz energy contribution due to hydrogen bonding

m^{seg}	Hard-chain segment number
σ	Hard-chain segment diameter
k_B	Boltzmann constant
u/k_B	Dispersion-energy parameter
ϵ^{AB}/k_B	Association-energy parameter
κ^{AB}	Association-volume parameter
n^{assoc}	Number of association (hydrogen bonding) sites
k_{ij}	Binary dispersion-energy correction parameter
K_a	Acid dissociation constant
$a(HA^{\pm})$	Activity of non-dissociated acid
$a(A^{-})$	Activity of acid after one dissociation step
$a(A^{2-})$	Activity of acid after two dissociation steps
$a(H_3O^{+})$	Activity of hydronium
$x^{L}_{HA,total}$	Total solid solubility of acid
$x^{L}_{HA,\pm}$	Solubility of non-dissociated acid
L', L''	Liquid phase designators
$x^{L,PC-SAFT}$	PC-SAFT modeled solid solubility
$x^{L,exp}$	Experimental solid solubility
$OF1$	Objective function for fitting pure-component parameters and binary interaction parameter between pure components and organic solvent
$OF2$	Objective function for fitting m^{seg}, σ, and u/k_B
NP	Number of experimental data points
$\rho^{PC-SAFT}$	PC-SAFT modeled density
ρ^{exp}	Experimental density
$P^{PC-SAFT}$	PC-SAFT modeled vapor pressure

P^{exp}	Experimental vapor pressure
AAD	Average absolute deviation
ARD	Average relative deviation

Mathematical Symbols Used in CHAPTER 5, Section 5.2 and APPENDIX C.

$\vdots\vdots\vdots^{c}$	Superscript indicates variable is reported on a concentration basis (except for pH)
$\vdots\vdots\vdots^{*}$	Superscript indicates variable is reported on a mixed-mode basis (except for pH)
$\vdots\vdots\vdots^{th}$	Superscript indicates variable is reported on a thermodynamic basis (except for pH)
$\vdots\vdots\vdots_{tot}$	Subscript indicates variable is totaled amount over all states
$\vdots\vdots\vdots_{init}$	Subscript indicates variable refers to an initial condition
$p\ \vdots\vdots\vdots$	Base-10 logarithm of the following variable
K_a	Acid association or stability constant (in general)
K_a^{γ}	Lumped activity correction
$\Delta \log K$, pK_a^{γ}	Base-10 logarithm of lumped activity correction
K_d	Acid dissociation constant (in general)
K_w	Auto-ionization constant of water (in general)
pK_a	Log of acid association constant (in general)
pH	Log of proton concentration or activity (in general)
pH_a	Log of proton activity
pH_c	Log of proton concentration
z	Charge of a conjugate base
j	Number of acidic protons
a	Thermodynamic activity
i	General indexing variable

I, IS Ionic Strength

γ Activity coefficient (in general)

ζ Charge of a single ion

r Reactant in a chemical reaction

p Product in a chemical reaction

ϕ, χ Lumped charge coefficient

A, B, C Empirical fit parameters, and hydrolysis reaction participants

Ind pH Indicator

X Reaction conversion

Mathematical Symbols Used in CHAPTER 5, Section 5.3:

$\vdots\vdots\vdots^{c}$ Superscript indicates variable is reported on a concentration basis (except for pH)

$\vdots\vdots\vdots^{th}$ Superscript indicates variable is reported on a thermodynamic basis (except for pH)

$\vdots\vdots\vdots_{tot}$ Subscript indicates variable is totaled amount over all states

$\vdots\vdots\vdots_{init}$ Subscript indicates variable refers to an initial condition

$\vdots\vdots\vdots_{\lambda_j}$ Subscript indicates variable is reported for the j^{th} wavelength

S Set of possible protonation states, i.e., $\{HA, A^-\}$ for a monoacid or $\{H_2A, HA^-, A^{2-}\}$

${}^{S}\vdots\vdots\vdots$ Preceding superscript indicates variable is reported for all protonation states in the set S

${}^{HA}\vdots\vdots\vdots$ Preceding superscript indicates variable is reported for a protonated weak monoacid

${}^{HA^-}\vdots\vdots\vdots$ Preceding superscript indicates variable is reported for a singly protonated weak diacid

${}^{A^-}\vdots\vdots\vdots$ Preceding superscript indicates variable is reported for a deprotonated weak monoacid

$^{A^{2-}}\vdots\vdots\vdots$	Preceding superscript indicates variable is reported for a fully deprotonated weak diacid
$\vdots\vdots\vdots^{pH_{c,i}}$	Superscript indicates that the variable is reported at the i^{th} pH_c value
$\vdots\vdots\vdots_{cell}$	Subscript indicates that the variable relates to the calorimeter cell
$\vdots\vdots\vdots_{syringe}$	Subscript indicates that the variable relates to the titrant syringe
$p\ \vdots\vdots\vdots$	Base-10 logarithm of the following variable
$\vdots\vdots\vdots_a$	Subscript indicates that the variable relates to the initial condition of a dilution
$\vdots\vdots\vdots_b$	Subscript indicates that the variable relates to the final condition of a dilution
i, j	General indexing variable
pH_c	Log of proton concentration
f	Fraction of a weak acid in the designated protonation state
pK_a	Log of acid association constant (in general)
λ	Wavelength, or eigenvalue
ℓ	Path length of optical cuvette
A	Analyte acid, or eigenvalue matrix, or empirical fit parameter
ε	Molar extinction coefficient
Q	Heating rate
V	Volume
C	Concentration, or empirical fit parameter
B	Buffer, or empirical fit parameter
K_a	Acid association or stability constant (in general)
K_w	Auto-ionization constant of water (in general)
K_a^γ	Lumped activity correction

α_i	Fraction of analyte acid in the i^{th} deprotonation state
β_i	Fraction of buffer in the i^{th} deprotonation state
x	Eigenvector
ξ	Molar extent of a neutralization reaction
n	Number of moles
$\Delta H_{neutralization}$	Total enthalpy of an injection due to neutralization reactions
ΔH	Molar enthalpy of a neutralization reaction
L_ϕ	Relative apparent partial molal enthalpy of infinite dilution
$L_{\phi,adjusted}$	Adjusted relative partial molal enthalpy of infinite dilution
m	Molal concentration
$\Delta H_{dilution}$	Molal enthalpy of dilution
$\Delta_{m_a}^{m_b} H_m$	Molal enthalpy of dilution from molality a to molality b
a, b, c, d	Empirical fit parameters
ω	Initial slope of a plot of adjusted relative partial molal enthalpy of dilution vs. the square root of molality
ΔH_{excess}	Excess molar enthalpy of mixing
$\mathscr{b}$	Binary interaction coefficient
AAD	Average absolute deviation
ARD	Average relative deviation
a	Thermodynamic activity

Mathematical Symbols Used in CHAPTER 6:

$\Delta G_{folding}$	Free energy of protein folding
$K_L a$	Oxygen mass-transport coefficient
ξ_1, ξ_2	Molar extent of reaction

LIST OF ABBREVIATIONS

2,3-DHBD_Ao	2,3-dihydroxybenzoate decarboxylase from *Aspergillus oryzae*
2,6-DHBD_Rs	2,6-dihydroxybenzoate decarboxylase from *Rhizobium sp.*
4-HZBA	4-hydrazonobenzoic acid
6-APA	6-aminopenicillanic acid
AAD	Average absolute deviation
AEH	Alpha amino acid ester hydrolase
APANiB	5-(2'-amino-5-phenylacetamido)-2-nitrobenzoic acid
ARD	Average relative deviation
AMF	5-Alkoxymethylfurfural
AUC	Analytical ultracentrifugation
β-LA	Beta lactam antibiotics
BTX	Benzene, toluene, xylene
CD	Circular dichroism spectroscopy
Cel	Cellulose, starch
CoA	Coenzyme A
CocE	Cocaine esterase
DFF	2,5-Diformylfuran
DHMF	2,5-Dihydroxymethylfuran
DKP	diketopiperazine
DMAP	Dimethylallyl-monophosphate
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide

DNTB	5-5’-dithiobis(2-nitrobenzoic acid), Ellman’s reagent
D-PG	D-phenyl glycine
D-PGME	D-phenyl glycine methyl ester
DSC	Differential scanning calorimetry
DSF	Differential scanning fluorimetry
EOS	Equation of state
ePC-SAFT	Electrolyte perturbed-chain statistical associating fluid theory
EtOH	Ethanol
FA	2-Furan carboxylic acid
FAD	Flavin adenine dinucleotide
Fdc1	Ferulic acid decarboxylase 1
FDCA	2,5-Furan dicarboxylic acid
FDME	2,5-Furan dicarboxylic acid methyl ester
FFA	5-formyl-2-furancarboxylic acid
FMN	Flavin mononucleotide
Fru	Fructose
FT-IR	Fourier transform infrared spectroscopy
Fur	Furfural
Glc	Glucose
Hem	Hemicellulose
HFCS	High fructose corn syrup
HMF	5-Hydroxymethylfurfural
HMFA	5-(Hydroxymethylfuran)-2-carboxylic acid
HmfF	FDCA decarboxylase 1
HmfG	Flavin prenyltransferase UbiX (formerly: FDCA decarboxylase 2)

HmfH	HMF/furfural oxidoreductase
HMFO	HMF oxidase
HPLC	High-performance liquid chromatography
HTP	High throughput
iBa	Isobutanol
ICP-OES	Inductively coupled plasma optical emission spectrometry
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Miller's LB broth; Luria-Bertani broth
LDA	Lithium diisopropylamide
LLE	Liquid-liquid equilibrium
Man	Mannose
MCS	Multiple cloning site
MEG	Monoethylene glycol
MOF	Metal organic framework
NE	Non-enzymatic
Ni-NTA	Nickel-nitrilotriacetic acid
PA	Polyamides
Pad1	Flavin prenyltransferase UbiX
PBM	Plant biomass
PCR	Polymerase chain reaction
PC-SAFT	Perturbed-chain statistical associating fluid theory
PDO	1,3-Propanediol
PEF	Poly(ethylene furanoate)
PET	Poly(ethylene terephthalate)

prFMN	Prenylated flavin mononucleotide
PTF	Poly(trimethylene furandicarboxylate)
PTT	Poly(trimethylene terephthalate)
PX	*p*-Xylene
QV	Quadruple variant
RCSB PDB	Research Collaboratory Structural Bioinformatics, Protein Databank
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SAD_Tm	Salicylic acid decarboxylase from *Trichosporon moniliiforme*
SAFT	Statistical associating fluid theory
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC-MALS	Size exclusion chromatography with multi-angle light scattering
SUMO	Small ubiquitin-like modifier
TA	Terephthalic acid
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TK	Transketolase
UbiD	3-octaprenyl-4-hydroxybenzoate decarboxylase
UbiX	Flavin prenyltransferase (formerly: propably aromatic acid decarboxylase)
UniProtKB	Universal Protein Resource Knowledgebase
WT	Wild-type
XRD	X-ray powder diffraction
Xyl	Xylose

SUMMARY

This research project explored the selection and development of an enzymatic route from the renewable feedstocks furfural and carbon dioxide to 2,5-furandicarboxylic acid, a building block for bio-based polyesters. A putative prenylated flavin-dependent reversible decarboxylase was identified as the most reasonable biocatalyst choice, although it was not isolated in a stable and active form in this work.

To support future reaction engineering endeavors, this project also included studies of the aqueous liquid and solid solubility relationships between the reactive pathway intermediates and possible side products: furfural, 2-furancarboxylic acid, 5-formyl-2-furancarboxylic acid, and 2,5-furandicarboxylic acid.

Finally, this work included an exploration into the mathematics of pH equilibria in aqueous reaction systems. These mathematical techniques were used to develop tools for assessing such systems by both colorimetric and calorimetric methods.

CHAPTER 1. INTRODUCTION

1.1 Motivating the Investigation of Routes to 2,5-Furandicarboxylic Acid

Rising levels of atmospheric carbon dioxide (CO_2) are generally accepted as the most significant anthropogenic contributor to radiative forcing and climate change. [1-2] As a result, carbon capture projects have gained traction in recent years, although the economics and acceptance of sequestering CO_2 into long-term storage are presently unclear. A plausible alternative is to use CO_2 as a feedstock for organic synthesis, which would also mitigate industrial dependence on petroleum-derived feeds. Emissions trading may permit some organizations to view carbon dioxide as a 'negative cost' input, further incentivizing the transition from petroleum. This tactic could yield beneficial changes in atmospheric CO_2 trends if widely adopted and incorporated into high-volume processes. [3] Even in the case that CO_2 cannot be used as a feedstock directly, the incorporation of plant biomass (PBM) into high-value chemical products could gradually serve to displace petroleum in favor of 'closed cycle' carbon as a chemical building block, and even help to support domestic growth of biorefineries whose primary focus is high-volume production of renewable fuels such as bioethanol and biodiesel. [4]

The manufacture of polyesters is a prime target for these CO_2-based and bio-based approaches, with worldwide annual production of poly(ethylene terephthalate) (PET) alone projected to exceed 21 million metric tons by 2021. [5] PET is produced by the direct esterification of terephthalic acid with excess ethylene glycol. [6] Terephthalic acid is predominantly made by selective oxidation of *p*-xylene, which in turn is produced by

catalytic reformation of petroleum naphtha. The Coca-Cola Company, which uses tremendous amounts of PET in packaging recently pledged to reduce the carbon footprint of its plastic packaging. At the start of this thesis work, the company was ramping up the use of bio-based ethylene glycol for PET polymerization for their 'Plantbottle' project. This new route, however, still yields a polymer that is 70 % derived from crude oil-sourced *p*-xylene. [7]

A similar polymer, poly(ethylene furanoate) (PEF), may be more amenable to production from biologically sourced carbon feeds. Unlike PET, which is built around the six-membered phenyl ring of terephthalic acid, PEF is built around the five-membered furan ring of the monomer 2,5-furandicarboxylic acid (FDCA). [8] A structural comparison of each monomer and its downstream polyester is shown in Figure 1.

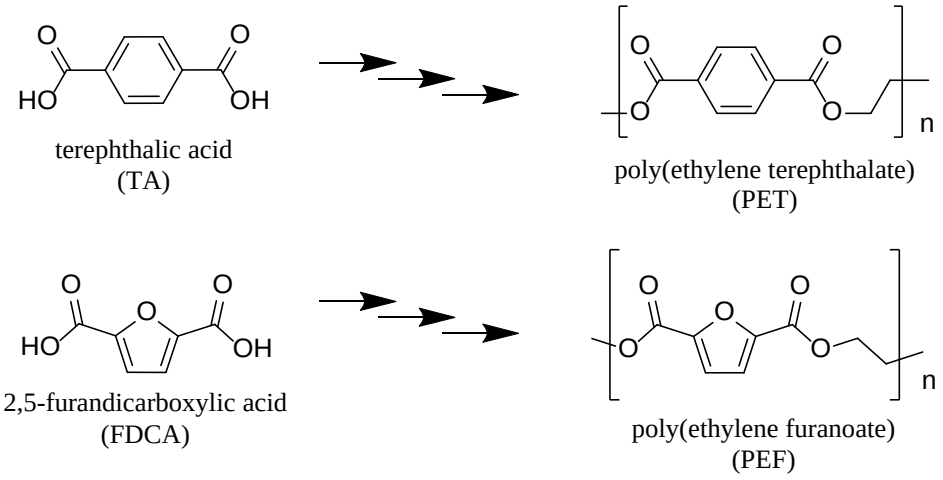
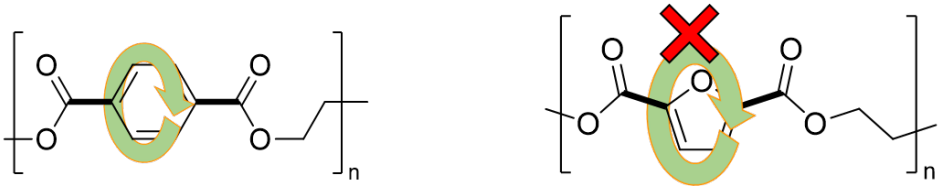


Figure 1: Visual comparison of monomers TA and FDCA, and their respective polymerization products, PET and PEF.

Due to the structural similarities between the two polymers, PEF promises to share in many of PET's desirable properties – it is clear, food-safe, and recyclable. The structural differences between PET and PEF, however, seem to lean in PEF's favor: the furan ring moiety makes FDCA (and thus ultimately PEF) ideally suited for production from furan-laden crop residues. [9] In fact, when bio-based FDCA is polymerized with bio-based ethylene glycol, the result is a wholly bio-based polyester. [7] Additionally, the bond angles on either side of the furan ring create a strain in the polyester that increases rigidity and resists 'ring flipping' (Figure 2), making PEF stronger, more elastic, and less permeable than PET to potential penetrants such as water, CO_2 and O_2. [10-11] Table 1 provides a comparison of some of the properties of PET and PEF. The promise of PEF to maintain or improve upon the desirable properties of PET suggests that it has the potential to displace PET in many industrial and consumer applications. [12]



PET PEF

Figure 2: Ring flipping in PET creates sweeping microvoids that contribute to the permeability of CO_2 and O_2. The bolded bonds indicate potential axes for ring flipping. PEF has bond angle strain in that is not present in PET. This strain increases polymer rigidity and reduces ring flipping, decreasing gas permeability and increasing tensile strength. Figure adapted from Burgess, et al. (2014) [10]

Table 1: Comparison of select physical and mechanical properties of PET and PEF.

Property	PET Value	PEF Value	Ref
Density ($g\ cm^{-3}$)	1.3346	1.4299	[10]
Glass Transition Temperature (°C)	76	85	[10]
Melting Temperature (°C)	247	211	[10]
Thermal Degradation Temperature (°C)	413	389	[10]
Diffusivity of $O_2 \times 10^{10}$ ($cm^2\ s^{-1}; at$ 35 °C)	97	10	[11]
Diffusivity of $CO_2 \times 10^{10}$ ($cm^2\ s^{-1}; at$ 35 °C)	22	0.72	[11]
Diffusivity of $H_2O \times 10^{10}$ ($cm^2\ s^{-1}; at$ 35 °C)	96	17	[11]
Storage Modulus (MPa; at 35 °C)	2120	3285	[10]
Tensile Modulus (GPa)	3.9-5.3	4.5-8.0	[13]
Yield Strength (MPa)	50-60	90-100	[13]
Max Tensile Strength (MPa)	170-250	120-300	[13]
Elongation at Break (%)	40-80	60-120	[13]
Quiescent Crystallization Time (min)	2-3	20-30	[13]

For these reasons, PEF and new routes toward its production have become increasingly popular areas of research in recent years. FDCA was number two on the list of "Top Value Added Chemicals from Biomass" in a 2004 report from the US Department of Energy also citing its potential for use in the production of succinic acid. [14] Figure 3 illustrates a growing trend in total publications and citations containing "2,5-furandicarboxylic acid" or "FDCA" over the last decade. Research for this thesis began in January of 2014.

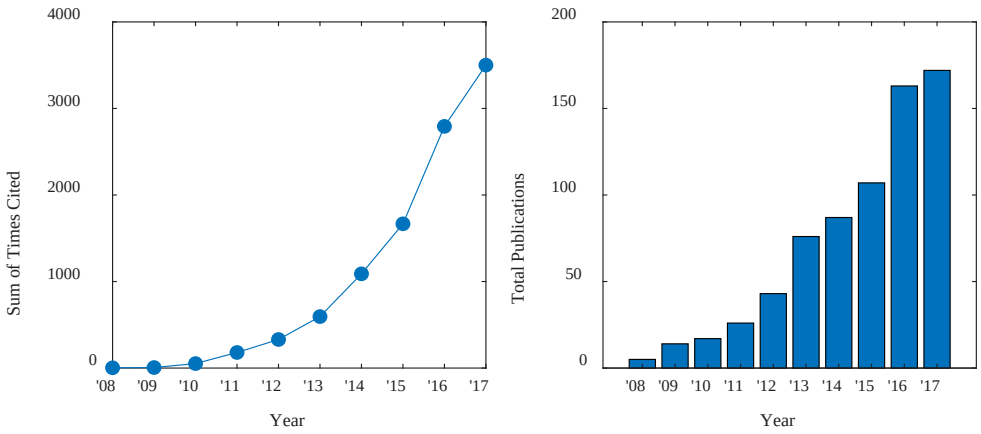


Figure 3: Trends in citations and publications containing either "2,5-furandicarboxylic acid" or "FDCA" from 2008 through 2017. Data are from Clarivate Analytics. [15]

1.2 A Survey of Routes to Renewable FDCA

Interest in FDCA has grown so rapidly, in fact, that even keeping track of the parties interested in its production, and their 'territory' in the intellectual property space is a challenge. The biopolymers 'playing field' has developed and shifted dramatically over the last decade, and even within the timeframe of this thesis. Figure 4 outlines a hypothetical network in which nodes represent chemical species and chemical processes are represented as directed edges (arrows). In this context, it should be noted that these network arrows are distinct from typical reaction arrows and are not intended to imply a balanced chemical reaction, or even to suggest that only a single reaction is involved in the jump from one node to another. The Euler diagram overlaid onto the network in Figure 4 illustrates how different organizations have divvied up some of the intellectual property in the biopolymer

space, based on a survey of publications, patents, patent applications, and corporate press releases. [16-37] This is just a small sample of the total biopolymer space, and does not even cover all the biopolymers derived from FDCA specifically. [38-41]

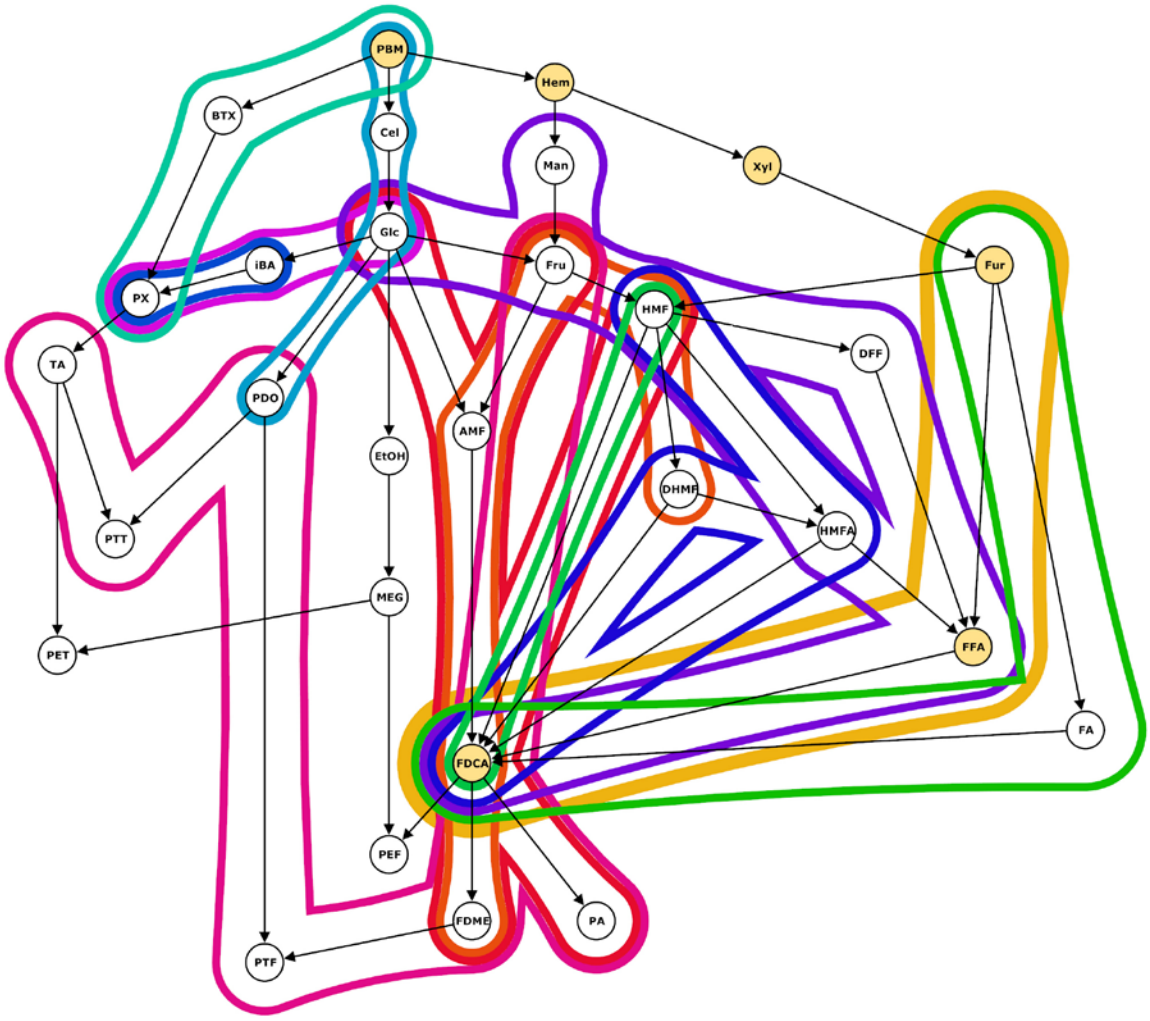
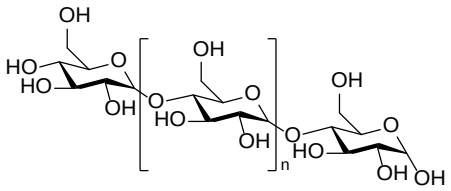
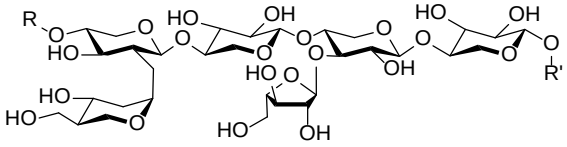


Figure 4: Some routes from PBM to biopolymers, and the organizations invested in their development. Molecular species are represented by the encircled nodes, and the connecting arrows represent the chemical processes that link them. The overlaid colored bands indicate which organizations have laid claim to each process: Georgia Institute of Technology (this thesis); Archer Daniels Midland Company; Furanix Technologies B.V.; E.I. du Pont de Nemours and Company; Gevo, Inc.; Novozymes; Purac Biochem B.V.; South Hampton Resources; Tate & Lyle plc; Virent; BASF SE; Stanford University. [16-37]

As Figure 4 suggests, there are many potential avenues from PBM to different bioproducts. Though there are some routes in development from PBM directly to BTX (benzene, toluene, and xylene), [29] the majority of the carbon-flux from biomass to valorized chemicals proceeds via cellulose and starch (Cel) and hemicellulose (Hem). While cellulose and starch hydrolyze to form glucose, the hydrolysis of hemicellulose yields a mixture of hexoses and pentoses including mannose (Man) and xylose (Xyl) (Figure 5).



Typical starch from food crops



Example hemicellulose

Figure 5: Structural comparison of starch and cellulose. Starch (left) has a regular structure and is comprised of glucose units. Hemicellulose (right) is a disordered polymer of many hexose and pentose sugars. [42]

Hexoses tend to be easily fermented, while pentoses are more recalcitrant (but not impervious) to biological conversion. [43] There is tremendous demand for hexoses from biomass for food products such as table sugar and high fructose corn syrup (HFCS). While there is some industrial demand for pentoses, e.g., for bioethanol or for production of the non-fermentable sweetener xylitol, there is much less demand than exists for hexoses. It is likely that anticipated competition for resources has played some part in driving the investigation of more routes to bioproducts from hemicellulose than from cellulose and

starch; which helps to explain why the right hand side of Figure 4 is so densely populated.[44-45]

What Figure 4 also helps to illustrate is that FDCA is a central hub for the production of biopolymers, leading not only to PEF, but also to poly(trimethylene furandicarboxylate) (PTF), and polyamides (PA). As a result of this centrality, as well as the attention brought to FDCA by the Department of Energy, there has been a technological race to develop and cordon off licensable technologies related to its manufacture.

The densely packed right lobe of Figure 4 is explored in more detail in Figure 6. Like Figure 4, Figure 6 outlines a hypothetical network in which nodes represent chemical species and chemical processes are represented as directed edges (arrows). Numbered network arrows, $\overset{i}{\rightarrow}$, are distinct from typical reaction arrows and are not intended to imply a balanced chemical reaction, or even to suggest that only a single reaction is involved in the jump from one node to another. The discussion below is divided between those routes that employ chemical catalysis and those that employ biocatalysis. Edges for which biocatalytic options are available are emphasized with green arrows, but there may be chemically catalyzed processes that correspond to these edges as well.

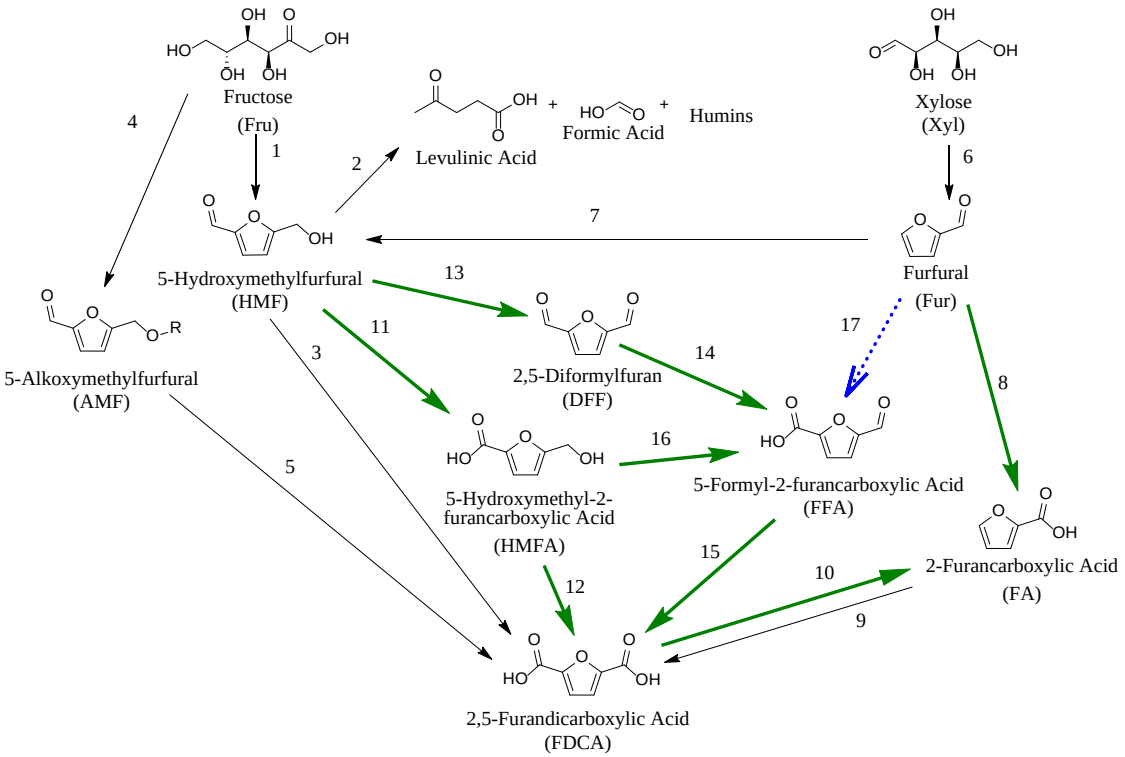


Figure 6: Comparison of chemical routes to FDCA from fructose and from xylose. Routes for which biocatalytic options are available are emphasized by green arrows. The numbered network edges are referenced in the text by this symbol: $\xrightarrow{i}$, where *i* is the edge number. Edge 17 (blue dashed arrow $\xrightarrow{17}$) has no prior art, and is explored in this thesis.

1.2.1 Routes to FDCA Employing Chemical Catalysis

As shown in edge 1 ($\xrightarrow{1}$ in shorthand**)**, 5-hydroxymethylfurfural (HMF) is primarily produced by the dehydration of hexoses, namely fructose. Dehydration is generally acid-catalyzed, although many reaction conditions have been explored, including the use of mineral acids or acidic ion exchange resins in aqueous and organic solvents, transition metal catalysts, and lanthanide(III) salts. [4, 9] A full review of HMF, including many routes

to its production was published in 2011.[46] HMF is unstable at high temperatures, and production yields following dehydration are typically limited by decomposition $\overset{2}{\rightarrow}$ to a mixture of levulinic acid, formic acid, and humins.[4, 9] These series-competing reactions and the general instability of HMF fundamentally detract from its merit as a chemical feedstock or a process intermediate. In fact, the US Department of Energy has stated that the poor selectivity of sugar dehydration processes, and the formation of unstable intermediates are the "primary technical barriers to production and use of FDCA."[14] Nevertheless, much collective effort has been invested in finding routes to FDCA from HMF. The multiple oxidation steps comprising $\overset{3}{\rightarrow}$ can be carried out under aqueous alkaline conditions with oxygen pressure and a PtPb catalyst with quantitative yield or in air with metal/bromide catalysts, though only 60 % yield is achieved.[4, 9] Sequential dehydration/cyclization and oxidation has been demonstrated to yield FDCA from fructose in a one-pot configuration ($\overset{1}{\rightarrow}$ and $\overset{3}{\rightarrow}$), in which the oxidation is performed in air and catalyzed by cobalt acetylacetonate in sol-gel silica, as well as in dimethylsulfoxide (DMSO) or dimethylformamide (DMF) using vanadyl phosphate catalysts with up to 97 % selectivity at 84 % conversion. The use of DMSO has been shown to significantly reduce levulinic acid formation.[4, 9]

Furanix Technologies B.V., a subsidiary of the Dutch company Avantium Technologies B.V., (which itself is a spin-off of Royal Dutch Shell) has been awarded several US and international patents relating to the synthesis of PEF from HMF. Citing the desire to produce fuel additives with high atom economy, one patent's claims include the

acid-catalyzed dehydration of glucose and fructose, with in-situ etherification of HMF with alcohols to provide modest yields of 5-alkoxymethylfurfural (AMF) ethers ($\overset{4}{\rightarrow}$). This concurrent etherification was found to reduce the decomposition of HMF described above and corresponding to $\overset{3}{\rightarrow}$. [26] A subsequent patent's claims include $\overset{5}{\rightarrow}$, the oxidation of these AMF species to FDCA in the presence of a subset of Co, Mn, and/or Br-based catalysts. [27, 32]

Alternative pathways to FDCA begin with $\overset{6}{\rightarrow}$, the dehydration of pentoses (namely xylose) in sulfuric acid at 200-250 °C, to yield furfural. [9] This process may also be performed in toluene/water mixtures at 160 °C using modified acidic zirconia catalysts; catalysts of heteropolyacids, titanate, niobate, and niobium silicate have also been explored for this purpose. [4] Furfural may be hydroxymethylated with aqueous formaldehyde over sulfonic ion-exchange resins ($\overset{7}{\rightarrow}$) using 1,3-dithiolate protection/deprotection on the aldehyde to push electron density into the furan ring. This reaction yields HMF with a selectivity of 90 %. [9]

Another option is to proceed along $\overset{8}{\rightarrow}$, the oxidation of furfural to 2-furancarboxylic acid (FA). This reaction has been demonstrated with 93 % selectivity at 100 % furfural conversion in aqueous phase under alkaline conditions at 60 °C, using air or oxygen as oxidant, with CuO nanoparticle catalysts. [47] This route has also been performed as a Cannizzaro reaction at temperatures less than 20 °C, with 60-63 % yields at 93-95 % purity, by NaOH-induced disproportionation to form sodium furan-2-carboxylate, and 2-furylcarbinol (not shown). The furylcarbinol is removed by extraction with ether and the

FA is protonated and crystallized by the addition of 40 % sulfuric acid, after which the remaining $NaHSO_4$ must be removed by further purification steps. [48]

The resulting FA could be converted to FDCA along the pathway indicated by $\overset{9}{\rightarrow}$. FA may be carboxylated in saturated NH_4Cl with lithium diisopropylamide (LDA) at 0 °C, or in tetrahydrofuran (THF) with LDA or n-butyllithium at -78 °C. In either case, the FDCA is extracted with ethyl acetate and HCl. The reaction is not regioselective, however, and also results in the formation of 2,3-furandicarboxylic acid (not shown). [49] Alternatively, a disproportionation reaction may be employed by adding $ZnCl_3$ to solid furoate (obtained after oxidation of furfural by removing water without acidification) to produce equimolar FDCA and furan. Optimal results were reached at 250-280 °C, using the sublimation of dry ice to increase the partial pressure of CO_2, for 86 % selectivity and 61 % conversion of the initial furfural. [47] More recently, FA under CO_2 atmosphere was carboxylated to FDCA in molten Cs_2CO_3.[16] This process was refined by the use of a Cs_2CO_3 - K_2CO_3 blend, which raised the melting temperature of the carbonate salts and enabled the use of a packed bed at 285 °C, resulting in molar scale production of FDCA with 89 % yield (139 g). [17]

1.2.2 Routes to FDCA Employing Biocatalysis

In 2010, Koopman and Wierckx reported their work isolating a strain of *Cupriavidus basilensis* capable of growth on minimal media using either furfural or HMF as the sole carbon source. [50] Species of the *Cupriavidus* genus are known for copper and heavy metal resistance, and for their diverse metabolic profiles that confer broad xenobiotic

degradation by chemoorganotrophic and facultative chemolithotrophic respiratory pathways. [51] In their HMF isolate, Koopman and Wierckx identified a metabolic pathway by which furfural and HMF are degraded prior to metabolic flux into the citric acid cycle. [52] Of principle interest to the present work are the putative pathway enzymes: 2,5-furandicarboxylic acid decarboxylase 1 (HmfF), flavin prenyltransferase UbiX (HmfG, previously named 2,5-furandicarboxylic acid decarboxylase 2), and HMF/furfural oxidoreductase (HmfH). Cell extracts containing the gene products of HmfF and HmfG together were found to decarboxylate FDCA to yield FA ($\overset{10}{\rightarrow}$). These activities are referenced in a 2011 patent application.[53] Furthermore, HmfH was found to perform $\overset{11}{\rightarrow}$, the oxidation of HMF to 5-hydroxymethyl-2-furancarboxylic acid (HMFA), as well as $\overset{12}{\rightarrow}$, the subsequent two-step oxidation of HMFA to FDCA. HmfH was also found to perform $\overset{8}{\rightarrow}$, the oxidation of furfural to FA. [54] *Gluconobacter oxydans* was also recently found to oxidize furfural to FA, but it was not evaluated for the conversion of 5-formyl-2-furancarboxylic acid (FFA) to FDCA. [55-56] *Acetobacter rancens* IFO3297, *Acetobacter pasteurianus* IFO13753, and *Serratia liquifaciens* LF14 were found to perform $\overset{8}{\rightarrow}$, $\overset{11}{\rightarrow}$, and $\overset{14}{\rightarrow}$ with high yield, but were found to not perform $\overset{15}{\rightarrow}$, the oxidation of FFA to FDCA. [57]

Attempts were made to oxidize HMF to FDCA using purified HmfH, but soluble expression could not be achieved. [4] More recently, however, HmfH and similar sequences were patented for the production of FDCA from HMF. [28] A soluble homologue to HmfH (46 % sequence identify) was identified in *Methylovorus* sp. strain MP688 and found to produce FDCA from HMF by sequentially traversing steps $\overset{13}{\rightarrow}$, $\overset{14}{\rightarrow}$, and $\overset{15}{\rightarrow}$ to yield (in

sequence) 2,5-diformylfuran (DFF), 5-formyl-2-furancarboxylic acid (FFA), and finally FDCA. This enzyme was then retrospectively nicknamed HMFO. [58] Though HMFO was found capable of converting HMFA to produce FFA ($\overset{16}{\rightarrow}$), HMFO was not found to produce HMFA as in $\overset{11}{\rightarrow}$ performed by HmfH, instead converting all HMF into DFF ($\overset{13}{\rightarrow}$). A 24 hour incubation of 20 μM HMFO with 4 mM HmfF and 20 μM flavin adenine dinucleotide (FAD) at ambient conditions was found to achieve 100 % substrate conversion, producing FDCA with 95 % yield. [59] The route traversing $\overset{11}{\rightarrow}\overset{16}{\rightarrow}\overset{15}{\rightarrow}$ (i.e., oxidation of HMF to HMFA, followed by dehydrogenase-catalyzed formation of FFA and subsequent oxidation to FDCA) has been described in a recent patent application. [60]

1.2.3 Heuristic Evaluation of Process Merits

While the preceding discussion illustrates that there are many possible routes from PBM to FDCA, it is evident that these pathways do not all represent choices that are desirable for 'green' industrial processes – particularly ones which could operate on a scale that might have a beneficial impact on atmospheric carbon. An ideal process would comply with green and sustainable chemistry guidelines, as well as generally accepted criteria for good chemical manufacturing processes. [61-62] Some simple process design heuristics can therefore be used to identify network edges or nodes with impractical attributes, and the graphs in Figure 4 and Figure 6 can be trimmed by the process of elimination.

For example: as the dehydration-loss of HMF typically occurs after the dehydration of fructose, all routes sharing $\overset{1}{\rightarrow}$ may be discarded as other routes exist which do not have

significant competing loss reactions, and will therefore have higher yield and atom economy. Similarly, the disproportionation reaction employed in $\overset{9}{\rightarrow}$ is eliminated, as half of all substrate goes to the production of furan. Rare, expensive, and toxic metal catalysts also incur penalties, as does the avoidable use of organic solvents; thus, the routes to furfural from xylose involving niobium and toluene are ruled out in favor of sulfuric acid dehydration. By similar reasoning, $\overset{3}{\rightarrow}$ and $\overset{5}{\rightarrow}$ are eliminated, in addition to pathways through $\overset{8}{\rightarrow}$ requiring either ether extraction or nanoparticles.

The resource competition mentioned previously is also an important consideration for process design. Historically, the first generation of biofuels, including those fermented from sugars engendered a significant increase in global demand for fermentable sugars. [44-45] If history is any indication, chemical routes developed for pentose feedstocks may be anticipated to induce less resource competition than those employing hexose feeds. In this light, routes from xylose appear much more favorable than those from fructose.

The incorporation of protection agents into a chemical process, such as the use of 1,3-dithiolate in $\overset{7}{\rightarrow}$, increases both the number of conversions required and the process mass intensity. The hydroxymethylation of furfural additionally requires the use of formaldehyde, which is a known human carcinogen based on established causal relationships between exposure and increased risks of nasopharyngeal, sinonasal, and lymphohematopoietic cancers in humans, including an analysis of over 11000 garment workers in Georgia and Pennsylvania. [63-64] Because PEF and related FDCA-based polyesters are likely to be used for food and beverage packaging and clothing, this route

should be avoided. Processes that require extremely low temperatures and pyrophoric reagents are both expensive and dangerous to operate, which rules out the routes through $\overset{9}{\rightarrow}$ requiring LDA or n-butyllithium.

Similarly, processes that require elevated temperatures are penalized if ambient temperature and pressure alternatives exist. Generally, this heuristic favors biocatalytic routes, but there is no path from xylose to FDCA that is traversable by biocatalytic options alone (even making allowances for the dehydration of xylose to furfural in sulfuric acid). In fact, the two edges that would enable a process (biocatalytic or otherwise) from furfural to FDCA are $\overset{7}{\rightarrow}$ and $\overset{9}{\rightarrow}$, which have been ruled out.

CHAPTER 2. PROPOSAL OF AN ALTERNATE BIOCATALYTIC ROUTE TO FDCA

This thesis was motivated by the apparent lack of a biocatalytic route from furfural to FDCA, and by the potential novelty of developing a reaction pathway incorporating inorganic carbon, CO_2, to directly upgrade pentose sugars by carboxylation. Were such a conversion possible, it would not only be "greener" than alternative process, but would also reduce competition for hexoses derived from crop residues that could otherwise be used in fermentation processes to produce biofuels. [43] The work described in the following sections is focused on the development of such a route, including the process of proposing a reaction scheme, and development of enzyme catalysts for those reactions. In support of anticipated biocatalyst optimization, reaction engineering, and process design tasks, some thermodynamic and physical property characterization was performed on furfural, FA, FFA, and FDCA, and tools were developed for evaluating and engineering the ionic equilibria of aqueous electrolyte systems, including for the development of high throughput colorimetric assays for enzyme development.

2.1 Chemical Route Selection

Two potential routes from furfural to FDCA were evaluated, both employing only biocatalysis. An enzymatic alternative to the chemically catalyzed $\overset{9}{\rightarrow}$ could be developed (perhaps by reversing the enzymatic decarboxylation activity of $\overset{10}{\rightarrow}$ with clever reaction engineering). This reversal would enable the pathway $\mathrm{Fur} \overset{8}{\rightarrow} \mathrm{FA} \overset{9}{\rightarrow} \mathrm{FDCA}$. An alternative

approach would be to develop an entirely new reaction from furfural to FFA (indicated by the dashed blue arrow, $\overset{17}{\rightarrow}$, in Figure 6). To date, the author is not aware of any prior art describing the direct carboxylation of furfural, and therefore developing such a reaction would open the pathway $\mathrm{Fur} \overset{17}{\rightarrow} \mathrm{FFA} \overset{15}{\rightarrow} \mathrm{FDCA}$, which is not otherwise possible.

It is not immediately apparent, however, whether either option is feasible, nor is it clear which of the two options is the most likely to lend itself to a successful manufacturing process. A simple thermodynamic estimate was employed to help address these questions – namely, which route would provide the highest equilibrium conversion, assuming that a perfect biocatalyst could eventually be developed. The standard free energy of formation, $\Delta_f G_i^\circ$, is the change in energy associated with forming one mole of a substance i in a defined standard state. The standard free energy of reaction, $\Delta_r G^0$, is a comparison of the standard free energies of formation of the species consumed or evolved by a reaction, according to Equation (1), where ν_i is the stoichiometric coefficient of species i.

$$\Delta_\mathrm{r} G^0 = \sum_i \nu_i \times \Delta_f G_i^0 \qquad (1)$$

Estimates of $\Delta_f G^0$ for the species participating in either route from furfural to FDCA were procured from the MetaCyc database of metabolic pathways within the BioCyc.org metabolic pathway database collection, and are provided in Table 2. [65-67] MetaCyc values are produced by estimating standard free energy of formation at pH equal to zero and ionic strength equal to zero according to a group contribution method. [68] Formation energies are then transposed to a pH of 7.3 using estimates of $\mathrm{pK_a}$ for

protonation equilibria, and adjusted to an ionic strength of 0.25 using an extended Debye-Hückel equation. [69] The standard error associated with these values is reported to be 2.22 kcal mol^{-1}, or about 9 kJ mol^{-1}. The MetaCyc database value for the free energy of formation of water differs substantially from that commonly associated with water at standard state, which was calculated using values from Cox, et al. (1989) [70] and is also presented in Table 2.

Table 2: Estimated standard free energies of formation of participants in the reactions from furfural to FDCA. Estimates are relative to standard a state defined by a pH of 7.3 and an ionic strength of 0.25. The data are collated from MetaCyc. [65-67] or (in the case of water) calculated from condensed-phase standard entropy and enthalpy of formation data. [70]

Compound	$\Delta_f G^0$ (kJ mol^{-1})
Furfural	-32 [a]
FA^-	-297 [a]
FFA^-	-422 [a]
$FDCA^{2-}$	-690 [a]
CO_2	-386 [a]
Bicarbonate	-545 [a]
H^+	2 [a]
H_2O_2	-48 [a]
O_2	16 [a]
H_2O	-151 [a]
	-237.065 [b]

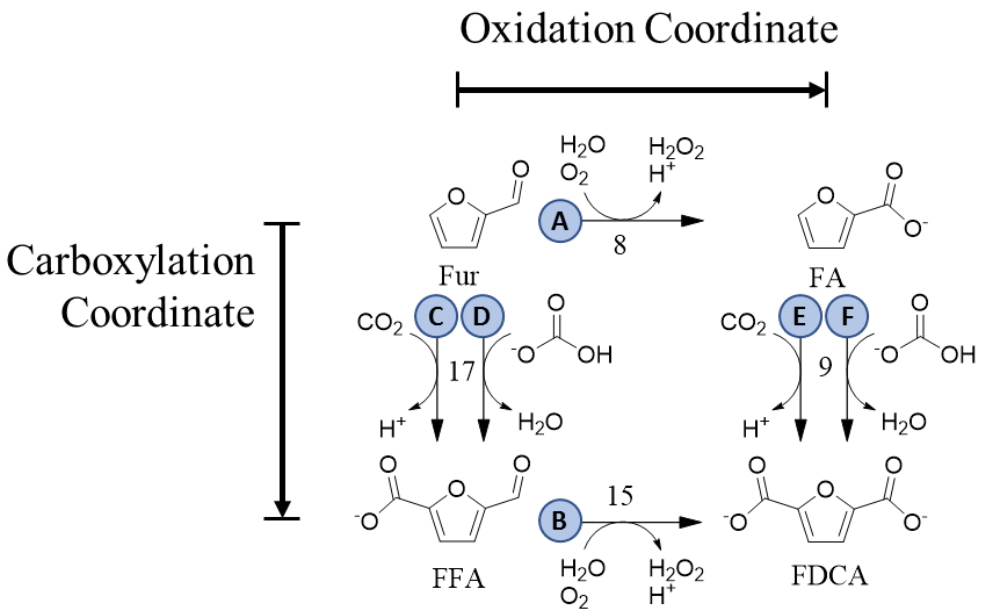
[a] Values from MetaCyc. [65-67]
Standard error is 9 kJ mol^{-1}

[b] Values calculated from Cox, et al. (1989) [70].
Standard error is neglected as it is very small relative to the other errors.

The values in Table 2 were used to estimate free energies of reaction for each step of either path from furfural to FDCA according to Equation (1), and the equilibrium constants, K_{eq}, for both carboxylation reactions were calculated using the relationship in Equation (2) at 25 °C, where R is the gas constant and T is the temperature in Kelvin. Reaction equilibrium constants are discussed in more detail in CHAPTER 4.

$$\Delta_r G^0 = -RT \ln\left(K_{eq}\right) \tag{2}$$

Uncertainty values were estimated using the rules for uncertainty propagation, starting from the standard error of $\Delta_f G^\circ$ estimates reported in Table 2. Note that the error propagation rules for antilog produce negative values for the lower uncertainty estimates of both carboxylation equilibrium constants, although the theoretical lower limit for an equilibrium constant is actually zero.



	MetaCyc Value Used for Water		Cox et al. Value Used for Water	
	$\Delta_r G^0$ (kJ mol^{-1})	K_{eq}	$\Delta_r G^0$ (kJ mol^{-1})	K_{eq}
A	-176 ± 23	$7 \times 10^{30} \pm 7 \times 10^{31}$	-90 ± 21	$7 \times 10^{15} \pm 5 \times 10^{16}$
B	-179 ± 23	$2 \times 10^{31} \pm 2 \times 10^{32}$	-93 ± 21	$2 \times 10^{16} \pm 2 \times 10^{17}$
C	-2 ± 19	2 ± 18	-2 ± 19	2 ± 18
D	4 ± 19	0 ± 2	-82 ± 16	$2 \times 10^{14} \pm 1 \times 10^{15}$
E	-5 ± 19	7 ± 52	-5 ± 19	7 ± 52
F	1 ± 19	1 ± 5	-85 ± 16	$7 \times 10^{14} \pm 4 \times 10^{15}$

Note: K_{eq} is mathematically bounded between zero and infinity.

Figure 7: Thermodynamic cycle comparing two routes from furfural to FDCA at pH 7.3 and an ionic strength of 0.25. Estimates of $\Delta_r G^0$ and K_{eq} are calculated from the values in Table 2. Alternate values are provided for the carboxylation reactions depending on whether carbon dioxide or bicarbonate is used as the inorganic carbon source for carboxylation, and whether the MetaCyc or Cox et al. values were used for the free energy of formation of water. The vastly different predictions for each approach underscore that this method is not sufficient for practical use in general. Estimated uncertainties are provided for each reaction. Reaction arrows are labeled with the number of the corresponding network edge in Figure 6. The route Fur $\xrightarrow{8}$ FA $\xrightarrow{9}$ FDCA ends in a thermodynamic trap, while the route Fur $\xrightarrow{17}$ FFA $\xrightarrow{15}$ FDCA benefits from continuous product removal from the equilibrium-limited carboxylation reaction.

On first inspection, both routes $\mathrm{Fur} \xrightarrow{8} \mathrm{FA} \xrightarrow{9} \mathrm{FDCA}$ and $\mathrm{Fur} \xrightarrow{17} \mathrm{FFA} \xrightarrow{15} \mathrm{FDCA}$ appear to be remarkably similar. Both routes start and end with the same compounds and therefore must have the same net standard free energy of reaction. The two routes each have both a direct carboxylation reaction and an oxidation step. Both routes also each have one established biocatalytic reaction ($\xrightarrow{8}$ and $\xrightarrow{15}$, respectively) and one that would need to be developed ($\xrightarrow{9}$ and $\xrightarrow{17}$). The vastly different estimates of $\Delta_r G^0$ and K_{eq} obtained using either the MetaCyc value for the free energy of formation of water, or the standard value of water calculated from Cox et al. only serve to underscore the overall uncertainty surrounding the prediction of equilibrium constants for aqueous reactions, and help to emphasize the fact that this method is, in general, not sufficient for practical use. Due to the large errors in estimation, both oxidation alternatives are essentially indistinguishable, but are likely favorable and irreversible, due to the large negative $\Delta_r G^{\circ}$. This prediction is consistent with the general knowledge that that oxidation reactions tend to be favorable, and is validated by the established work demonstrating that oxidation of FFA to FDCA by HMFO has been shown to proceed with nearly quantitative conversion and yield.[59, 71] The estimated uncertainties for the standard free energies of the carboxylation reactions are large relative to the actual calculated $\Delta_r G^0$ prediction. Because the uncertainties of these values span zero, the predicted equilibrium constants span values above and below unity. This prediction is consistent with the observation that carboxylation reactions are typically thermodynamically limited due to the high oxidation state and low energy of CO_2, as well as experimental demonstrations of the reversibility of many enzymatic decarboxylation reactions (which are discussed in more detail in CHAPTER 3, section 3.1.1).[72-79]

In fact, the only clear difference between the two routes in Figure 7 is the relative ordering of the carboxylation and oxidation steps. In the route that passes via FFA, carboxylation is the first step, while it is the second step in the alternate route proceeding via FA. In this light then, the choice between the two routes is really a question of whether an equilibrium-limited reaction should come before or after an irreversible one: $A \rightleftharpoons X \rightarrow B$, or $A \rightarrow Y \rightleftharpoons B$. If the equilibrium limited reaction occurs first in series, then the irreversible reaction lends itself to a form of product-removal. Whatever limited intermediate is formed will be irreversibly reacted away to the final product, and the reaction quotient will drop, driving greater conversion of the primary reactant. The equilibrium limit of this hypothetical batch reactor lies at or close to complete conversion of reactants to products. Unsurprisingly, the alternate route presents a very different scenario: complete conversion of the primary reactant into the intermediate, followed by equilibrium limited conversion to final product. The overall equilibrium of this hypothetical batch reactor is still limited by the reversible reaction, and appreciable conversion would not be possible without in-situ product removal, such as by continuously precipitating the final product. While reactive crystallization systems do exist, they are difficult to implement – and this is especially true when biocatalysts are involved. [80-81] For this reason, the route $\mathrm{Fur} \xrightarrow{17} \mathrm{FFA} \xrightarrow{15} \mathrm{FDCA}$, with potentially-equilibrium-limited carboxylation followed in sequence by irreversible oxidation was selected as the preferred option by which to proceed.

CHAPTER 3. BIOCATALYST DEVELOPMENT

3.1 Biocatalyst Candidate Selection

3.1.1 Carboxylation Catalyst Candidate Selection

As of this writing, there is no prior art describing the direct carboxylation of furfural (the first reaction in Figure 8, and $\overset{17}{\rightarrow}$). To develop this reaction, a survey was performed to evaluate enzymes from the literature for use as template biocatalysts which could later be improved upon with an assortment of protein engineering tools.

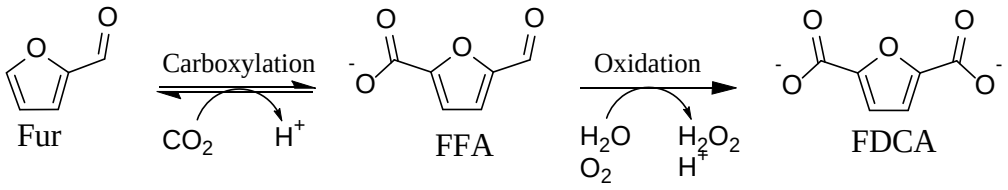


Figure 8: Proposed series reaction scheme for the conversion of furfural to FDCA. Furfural is reversibly carboxylated to FFA, likely under sufficiently basic pH that the resultant carboxylic acid group is deprotonated. This reaction is followed in series by the irreversible oxidation to FDCA, converting the aldehyde on FFA to a second carboxylic acid group that is likely also deprotonated. Due to the highly negative $\Delta_r G^\circ$, the oxidation reaction will likely continue to completion, performing the dual role of product removal for the reversible reaction, allowing complete conversion of furfural to FDCA.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) received substantial press for catalysing the addition of CO_2 (on the order of 10^{11} metric tons per year) to ribulose 1,5-bisphosphate in the Calvin-Benson-Bassham cycle. Though it has been studied

continuously since 1947, it has yet to see any industrial synthesis applications outside of living photosynthetic tissue. [82] RuBisCO, however, is not the only enzyme capable of catalyzing the addition of CO_2 to anabolize organic molecules. In fact, many biological carbon fixation pathways exist, though all apparently require substantial energy input to overcome the high oxidation state of CO_2.

Many naturally occurring aromatic decarboxylation reactions, however, are reversible and the associated biocatalysts can perform carboxylation under right conditions *in vitro*. [72] Phenols and hydroxystyrene derivatives have been regioselectively carboxylated by several "decarboxylase" enzymes, including benzoic acid decarboxylases and a salicylic acid decarboxylase. [73-76] Pyrrole-2-carboxylate decarboxylase from *Bacillus megaterium* PYR2910 was also shown regioselectively to carboxylate its natural product in the presence of concentrated carboxylate salts. [77-79] Three of the reversible aromatic decarboxylases evaluated by Wuensch, et al. (2012) [75] were selected for investigation as furfural carboxylases: 2,3-dihydroxybenzoate decarboxylase from *Aspergillus oryzae* (2,3-DHBD_Ao), 2,6-dihydroxybenzoate decarboxylase from *Rhizobium sp.* (2,6-DHBD_Rs), and salicylic acid decarboxylase from *Trichosporon moniliiforme* (SAD_Tm).

Generally speaking, it can also be useful to look for biocatalyst enzyme templates in metabolic pathways involving either the exact chemical participants in the desired reaction or structural analogues of those compounds. Furfural and HMF are toxic to many microorganisms, inhibiting the microbial fermentation of lignocellulosic hydrolysates by damaging DNA, inhibiting glycolysis, and disrupting the incorporation of sulfur into cysteine and methionine. [83] Ostensibly in response to selection pressure due to this toxicity,

several organisms have developed degradation pathways capable of eliminating furan-series compounds. *Pseudomonas* F2 isolated by enrichment culture was found to grow on minimal media with FA as the sole carbon source, forming a thioester linkage with coenzyme A (CoA) for shuttling through further degradation by a series of hydrolase enzymes to 2-oxoglutarate, which is injected into the TCA cycle (Figure 9). The *Pseudomonas* isolates were found not to grow on furfural. [84]

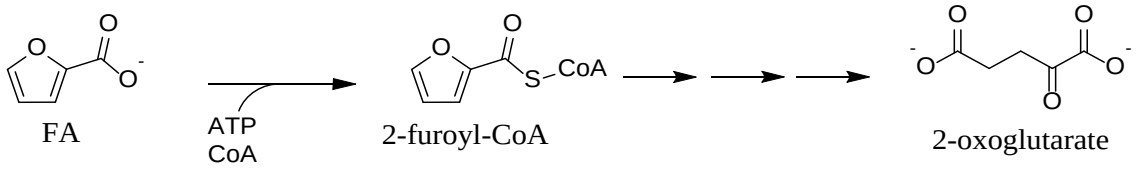


Figure 9: Degradation of FA leading to the TCA cycle. FA is tagged by thioester linkage to CoA, and iteratively hydrolyzed to yield 2-oxoglutarate which is used by the TCA cycle. [84]

Similar work conducted more recently determined that *Cupriavidus basilensis* HMF14 could grow on minimal media with either furfural or HMF as the sole carbon source. Three enzymes discovered in this pathway are of principle interest to this work. HMF/furfural oxidoreductase (HmfH) was found to oxidize HMF to FDCA, and to oxidize furfural to FA. Collectively, 2,5-furandicarboxylic acid decarboxylase 1 (HmfF) flavin prenyltransferase UbiX (HmfG, previously named 2,5-furandicarboxylic acid decarboxylase 2) were found to decarboxylate FDCA to FA, as shown in Figure 10, and discussed in the introduction section 1.2.2, Routes to FDCA Employing Biocatalysis. A

similar pathway to that described for *Pseudomonas* F2 converts the FA to 2-oxoglutarate for use in the TCA cycle. [52-54, 85-87]

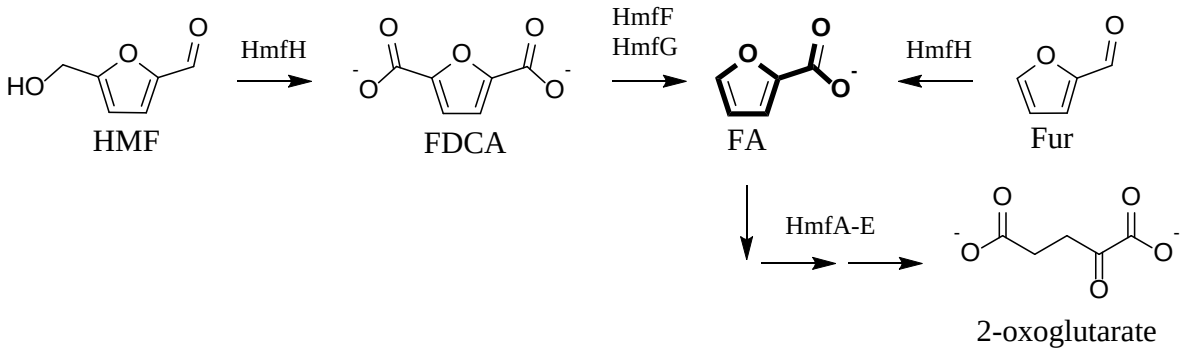


Figure 10: Degradation of HMF, FDCA, FA, and Furfural by *C. basilensis* HMF14. HmfH oxidizes HMF to FDCA and furfural to FA. Collectively, HmfF and HmfG decarboxylate FDCA to FA. FA is converted to 2-oxoglutarate which is used in the TCA cycle. [52-54, 85-87]

A more detailed pathway map with gene and enzyme annotations, as well as links to other established metabolic pathways is available on the Kyoto Encyclopedia of Genes and Genomes (KEGG). [85-87] The curated pathway for the degradation of furfural and HMF, map00365, is reprinted with permission in Figure 11. A similar map is available on MetaCyc. [65-67] Pathway maps such as these are invaluable tools for identifying new opportunities for free-enzyme biocatalysts and for the metabolic engineering necessary to develop whole-cell biocatalysts.

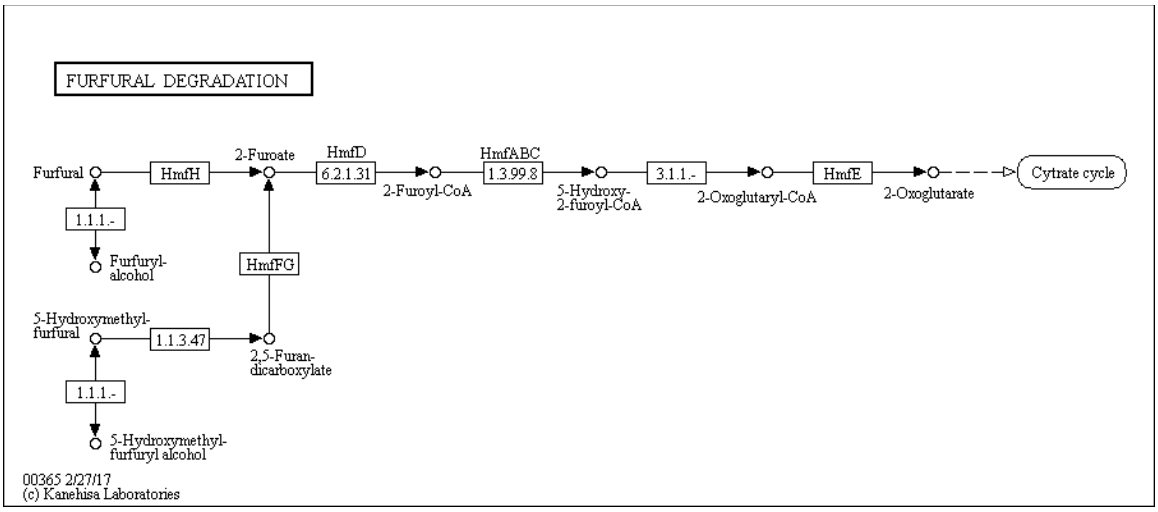


Figure 11: Furfural degradation pathway map from KEGG. Database-curated metabolic pathways are invaluable resources for biocatalyst template selection. Pathway map00365 reprinted with permission. [85-87]

Based on the apparent equilibrium limit of furfural carboxylation (Figure 7) and the reversibility demonstrated by other carboxylic acid decarboxylases mentioned previously, it was assumed that the decarboxylation reaction catalyzed by the HmfF/HmfG pair is also reversible such that the duo would likely be able to catalyze the carboxylation of FA to FDCA. FA is structurally similar to furfural, so it was also assumed that the HmfF/HmfG pair would be reasonably likely to catalyze the carboxylation of furfural to FFA. The HmfF/HmfG pair were also selected for investigation as biocatalysts templates.

When research towards this thesis began, both HmfF and HmfG were putatively classified as FDCA decarboxylases. All that had been established about their behavior was that HmfF was necessary to imbue cell extracts with the ability to decarboxylate FDCA, and that this activity was significantly enhanced by coexpression with HmfG. Cell extracts

containing HmfG alone were not found to have any decarboxylation activity. [52] It was inferred from this observation that the HmfF/HmfG system is behaviorally similar to the previously established UbiD/UbiX type decarboxylase system. UbiX-deficient *Salmonella typhimurium* was known to be phenotypically similar to UbiD-deficient *Escherichia coli* due to their mutual inability to decarboxylate 3-polyprenyl-4-hydrodroxybenzoate to polyprenylphenol in the ubiquinone biosynthetic pathway (Figure 12). [85, 88-90] At that time, however, there was no evidence to suggest specifically what role each enzyme might play, or how they might interact to accomplish catalysis.

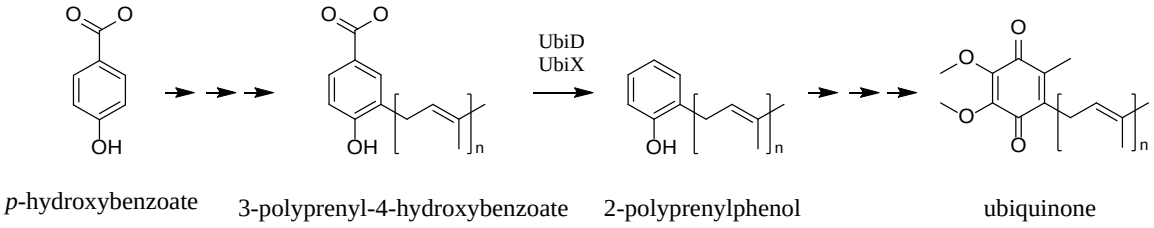


Figure 12: Ubiquinone biosynthesis pathway. HmfF/HmfG homologues UbiD and UbiX were known to be important for the decarboxylation of 3-polyprenyl-4-hydroxbenzoate in the ubiquinone biosynthetic pathway. UbiD was demonstrated to be necessary for decarboxylation. UbiX was known to have some supporting role in this process, but is not sufficient for decarboxylation in the absence of UbiD. [85-90]

In June of 2015, however, two publications revealed the specific structures, roles, and mechanisms of bacterial UbiD and UbiX, as well as the fungal homologues Fdc1 (ferulic acid decarboxylase 1) and Pad1 (previously phenylacrylic acid decarboxylase 1, but now called flavin prenyltransferase UbiX). [91-92] In these articles, it was proposed that the role of UbiX is to modify flavin mononucleotide (FMN) by the addition of a prenyl

group from dimethylallyl-monophosphate (DMAP) to yield an enzyme cofactor that had not previously been described – a prenylated FMN (prFMN). [91] This reaction is illustrated in Figure 13.

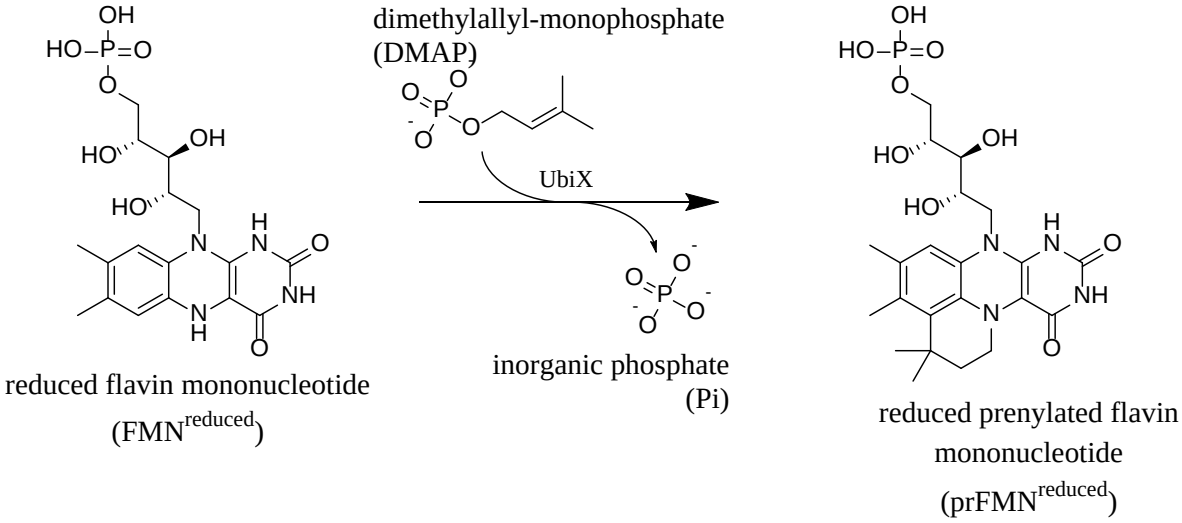
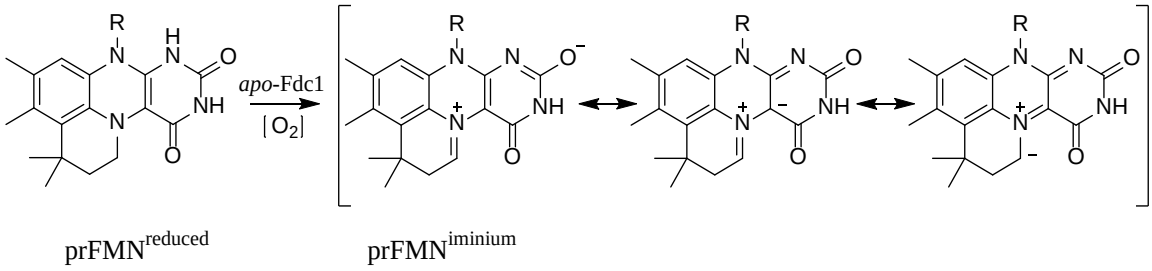


Figure 13: Prenylation of reduced FMN by UbiX. UbiX catalyzes the addition of a prenyl group to the N5 and C6 positions of the isoalloxazine ring of flavin mononucleotide, producing a reduced prenylated flavin (prFMN^{reduced}) and inorganic phosphate. [91]

It was subsequently demonstrated that the reduced prenylated flavin (prFMN^{reduced}) incorporates into the *apo* form of UbiD or Fdc1 where it oxidizes to a reactive N5-iminium (prFMN^{iminium}). The iminium form of the prenylated cofactor is proposed to exhibit azomethine ylide resonance, which is not available to unmodified flavin (Figure 14). This electron delocalization generates a dipole moment that may play a role in substrate alignment within the decarboxylase active site and is critical to the proposed decarboxylation mechanism. [92-94]



prFMN^{reduced} prFMN^{iminium}

Figure 14: Reduced prenylated FMN is taken up by Fdc1 and oxidized to the N5-iminium form. The iminium form, prFMN^{iminium}, performs the role of a resonant enzyme cofactor and exhibits azomethine ylide electron delocalization that is proposed to play an important role in substrate alignment and decarboxylation. [92-94]

The discovery of prFMN changed the landscape of research around biocatalytic reversible decarboxylation. The work has been found to explain much of the observed behaviors of the UbiX/UbiD family, as well as the ways in which members of the overarching family of reversible decarboxylases may differ. [95] Details of the catalytic mechanisms reliant on this new cofactor and the conserved structural features associated with decarboxylation are still being explored. [96] The initial publications detailing the discovery of prFMN had the broad effect of redefining many putative decarboxylases as putative flavin prenyltransferases. One of the enzymes that had already been selected for this thesis project, HmfG, was also a part of this sweeping reclassification. It was reported in a recent article discussing the new discoveries in relation to the family of decarboxylase enzymes that HmfF is likely a member of a family of hetero-aromatic acid decarboxylases. Another suggested member of this family is the pyrrole-2-carboxylate decarboxylase discussed previously, which is proposed to perform acid-base catalysis and is not suspected of using prFMN. However, pyrrole-2-carboxylate decarboxylase was never sequenced or

cloned and direct sequence comparison or phylogenetic analysis is therefore not currently possible.

3.1.2 Oxidation Catalyst Candidate Selection

A homologue to HmfH, HMFO, was discovered that is also capable of oxidizing HMF to FDCA via FFA as an intermediate. Unlike HmfH, which will oxidize furfural to FA, HMFO has no appreciable activity on furfural (Figure 15). Ligand-docking simulations and alanine-scanning were used to identify catalytically significant residues in the HMFO active site. Site-directed mutagenesis of these key residues yielded a double-variant, HMFO-V367R-W466F with 1000-fold higher catalytic efficiency than the wildtype. [52, 58-59, 71]

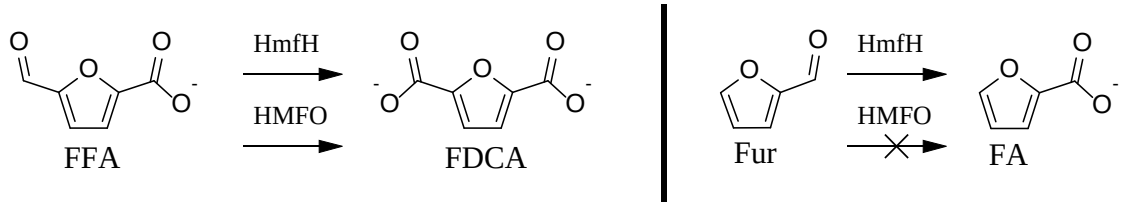


Figure 15: HmfH has activity on both FFA and Furfural, while HMFO has no activity on furfural. [52, 58-59, 71]

The chemical route that was ultimately selected for furfural conversion (Figure 8) requires that both carboxylation of furfural and subsequent oxidation of FFA occur in the same reaction vessel. This is a necessary condition to link the two reactions and realize the benefit of "product removal" imparted by the oxidation reaction. It follows, then, that the FFA oxidizing enzyme will come into direct contact with furfural. If HmfH is used to

catalyze the second reaction in Figure 8, then it will also oxidize furfural (as in Figure 10 and Figure 15), leading to the accumulation of FA, and dead-ending the reactor in the thermodynamic trap discussed in the preceding section and illustrated in Figure 7. In contrast, selection of HMFO as the biocatalyst for FFA oxidation precludes the formation of FA and eliminates the thermodynamic trap. To the best of this author's knowledge, HMFO is the only biocatalyst for which this substrate selectivity has been demonstrated. Therefore, the engineered double-variant HMFO-V367R-W466F was selected to be the oxidizing biocatalyst for future study.

3.1.3 Summary of Biocatalyst Candidate Selection Results

The enzymes selected for evaluation are listed in Table 3. Their protein sequences are provided in the appendix section A.1 Protein Amino Acid Sequences.

Table 3: Enzymes selected for evaluation.

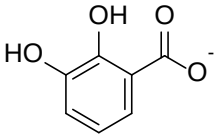
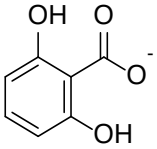
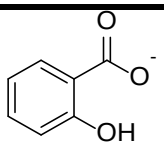
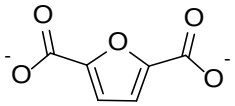
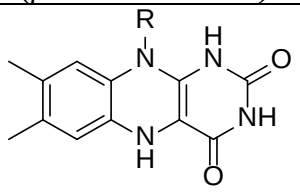
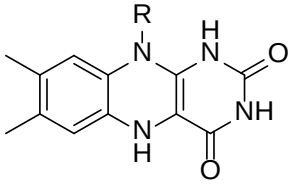
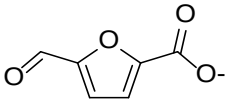
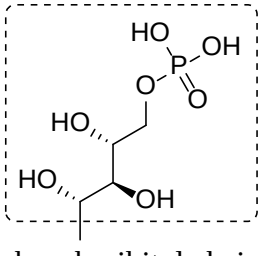
Desired Activity	Enzyme Name (and Nickname)	Primary Native Substrate	Ref
Carboxylation of furfural	2,3-Dihydroxybenzoic acid decarboxylase (2,3-DHBD_Ao)	 2,3-dihydroxybenzoate	75
Carboxylation of furfural	2,6-Dihydroxybenzoic acid decarboxylase (2,6-DHBD_Rs)	 2,6-dihydroxybenzoate	75

Table 3 (continued)

Desired Activity	Enzyme Name (and Nickname)	Primary Native Substrate	Ref
Carboxylation of furfural	Salicylic acid decarboxylase (SAD_Tm)	 salicylate	75
Carboxylation of furfural	2,5-furancarboxylic acid decarboxylase 1 (HmfF)	 2,5-furandicarboxylate *(putative substrate)*	52
Prenylation of FMN	Flavin prenyltransferase UbiX (previously named 2,5-furandicarboxylic acid decarboxylase 2) (HmfG)	 flavin mononucleotide *(putative substrate)*	52
Prenylation of FMN	Flavin prenyltransferase UbiX (UbiX)	 flavin mononucleotide	91
Oxidation of FFA	HMF oxidase V367R-W466F (HMFO)	 5-formyl-2-furancarboxylate *(engineered substrate)*	71
R =  phosphoribityl chain			

3.2 Structural Predictions and Sequence Alignments of HmfF and HmfG

3.2.1 Abstract

Sequence alignments and structural predictions were performed to investigate the homology between the HmfF/HmfG enzyme system and the Fdc1/UbiX enzyme system. Sequence alignments show that all residues in Fdc1 and UbiX for which functions have been assigned are conserved within their counterparts from *Cupriavidus basilensis*. Structural alignments between crystal structures of Fdc1 or Ubix and the homology models of HmfF or HmfG show that these conserved functional residues likely occupy similar positions within each enzyme.

3.2.2 Introduction

After the discovery of prFMN was published,[91-92] sequence and structural alignments were performed to evaluate whether the functions identified for Fdc1 and UbiX might also be expected for HmfF and HmfG. As of this writing, no crystal structures have been solved for either HmfF or HmfG from *Cuprivaidus basilensis* HMF14, so structural predictions were made to facilitate structural alignment. Crystal structures of two other enzymes with closer sequence homology to HmfF and HmfG than the Fdc1 and UbiX alignment targets were used as templates for the structural predictions.

3.2.3 Methods

Homology structures were generated using SWISS-MODEL to help provide some familiarity with the possible structures of the enzymes and to determine of residues

considered critical to the function of Fdc1 and UbiX were likely to occupy the same position in HmfF and HmfG.[97] Protein sequences were obtained from the Universal Protein Resource Knowledgebase (UniProtKB) database.[98] Crystal structures for Fdc1 and UbiX were obtained from the Research Collaboratory for Structural Bioinformatics Protein Databank (RCSB PDB).[99] The UniProtKB and RCSB accession identifiers for each enzyme (including those as structural templates for the homology models) are listed in Table 4. Additional protein sequence alignments were performed in CLC Sequence Viewer 7.7.1 (Qiagen, Redwood City, California).

Table 4: UniProtKB and RCSB PDB Accession Numbers for enzyme structure and sequence alignments.

Enzyme	Organism	UniProtKB Accession	RCSB Accession	Note
HmfF	*Cupriavidus basilensis*	D5KB58	NA	No structure available
HmfG	*Cupriavidus basilensis*	D5KB59	NA	No structure available
Fdc1	*Aspergillus niger*	A2QHE5	4ZA4	See ref. [92]
UbiX	*Pseudomonas aeruginosa*	Q9HX08	4ZAZ	See ref. [91]
ubiD	*Escherichia coli*	P0AAB4	2IDB	Template for HmfF
ecdB	*Escherichia coli*	P69772	1SBZ	Template for HmfG
NA: Not Applicable				

3.2.4 Results

3.2.4.1 HmfF Structural Prediction and Alignment with Fdc1

A UbiD from *E. coli* was used as the structural template for the HmfF homology model, as listed in Table 4. This template was selected by the HHblits algorithm based on a sequence identity of 30.29 %, a sequence similarity of 0.34, and a coverage of 0.90. [100] The resulting model was aligned with the crystal structure of Fdc1 in Pymol. The structural alignment between the enzymes was used to position the prenylated flavin ligand from the Fdc1 structure into the pocket of the HmfF homology model. [101] The two structures are placed side-by-side for visual comparison in Figure 16 and Figure 17.

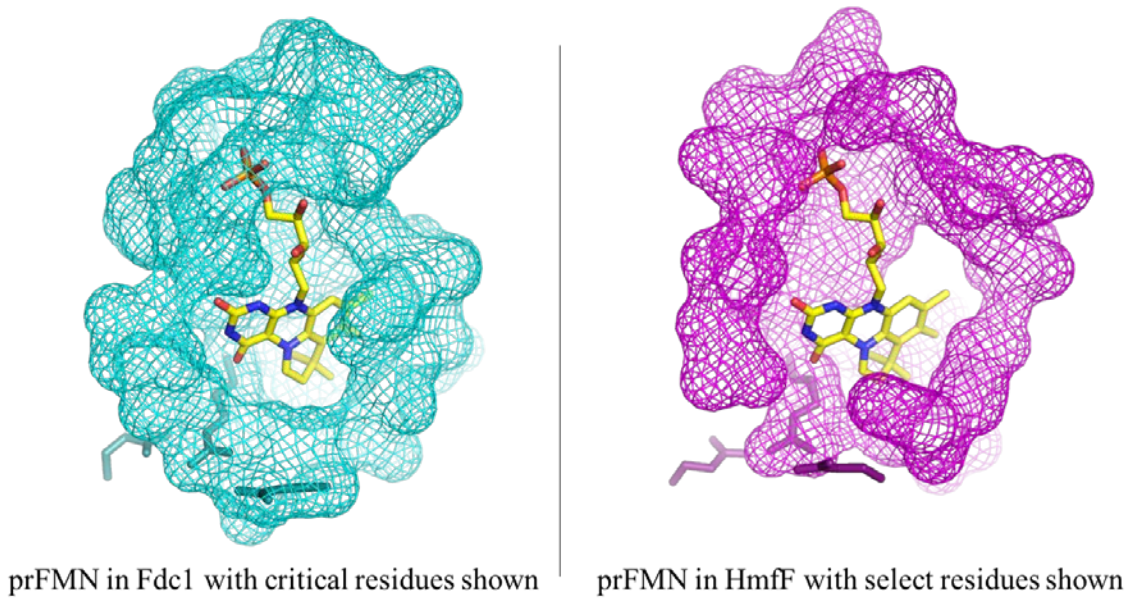


Figure 16: Side-by-side comparison of Fdc1 and HmfF with selected conserved residues shown. The crystal structure of Fdc1 is on the left, and the homology model developed for HmfF is on the right. Three important functional residues in Fdc1 near the head of the prenylated flavin are displayed for emphasis: (from left to right) E277, R173, and E282. These three residues are also present in the same general locations in the homology structure of HmfF.

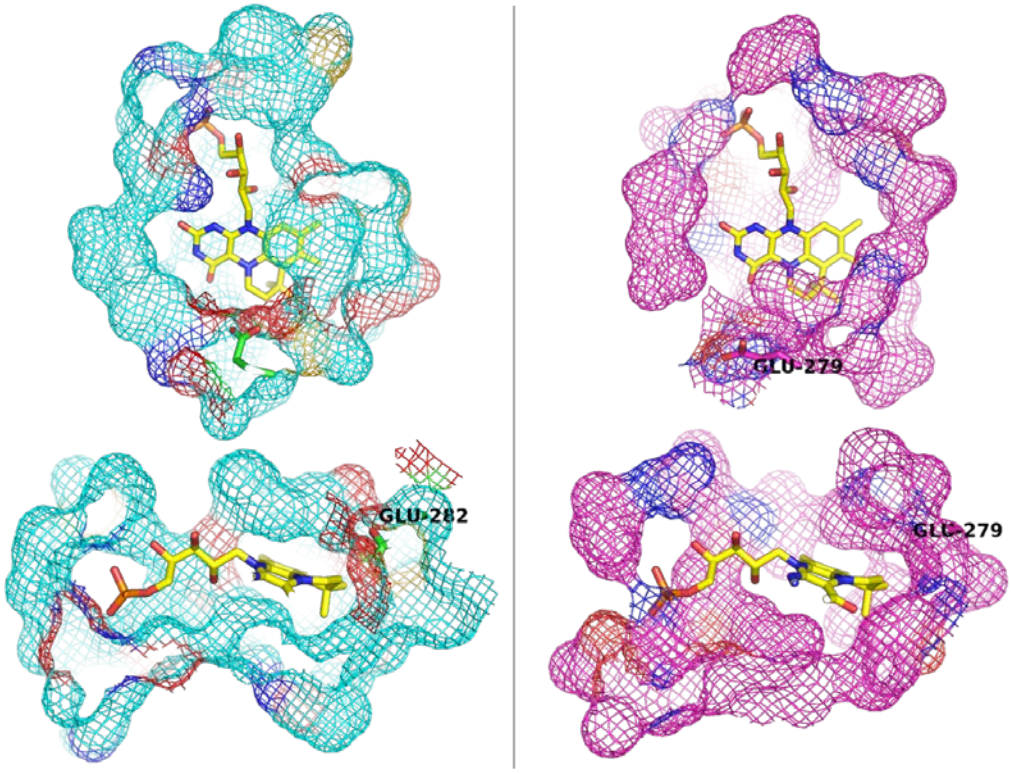


Figure 17: Side-by-side comparison of the binding pockets in Fdc1 and HmfF. The crystal structure of Fdc1 is shown on the left and the homology model of HmfF is shown on the right. The conserved catalytic glutamic acid residue is shown to help illustrate relative positioning.

The constituents of the polar E277-R173-E282 network in Fdc1 (Figure 16) has been shown to play key roles in the oxidative maturation of the prenylated flavin cofactor, and in subsequent catalysis of decarboxylation. Both E277 and E282 participate in the proposed decarboxylation mechanism of Fdc1, and their absence or substitution for non-acidic residues yields inactive enzyme. R173 is proposed to play a role in substrate orientation and has been demonstrated to be necessary for maturation of the reduced

prenylated flavin to the prFMN^{iminium} form. [92, 96] As Figure 16 shows, these residues are conserved within HmfF, and occupy the same positions relative to the cofactor, after structural alignment. Overall, HmfF was found to have 26.37 % sequence identity with Fdc1, with a sequence similarity of 0.33, and coverage of 0.92, based on the HHblits algorithm. [100] Of seven residues proposed to have significance to the functionality of Fdc1, six are identical in HmfF and one is similar. These residues and their proposed functions are shown in Table 5.

Table 5: Residues that are critical to Fdc1 function are conserved in HmfF.

Residue in Fdc1	Proposed Function	Conservation in HmfF
N168	Mn-dependent cofactor binding (tail)	Identical → N167
R173	Cofactor maturation	Identical → R172
Q190	Cofactor binding (head)	Similar → R188
H191	Mn-dependent cofactor binding (tail)	Identical → H189
E233	Mn-dependent cofactor binding (tail)	Identical → E230
E277	Catalysis	Identical → E274
E282	Catalysis	Identical → E279

3.2.4.2 HmfG Structural Prediction and Alignment with UbiX

A putative UbiX-like flavin prenyltransferase from *E. coli* was used as the structural template for the HmfG homology model, as listed in Table 4. This template was selected by the HHblits algorithm based on a sequence identity of 49.74 %, a sequence similarity of 0.43, and a coverage of 0.81. [100] The resulting model was aligned with the crystal structure of UbiX in Pymol. The structural alignment between the enzymes was used to

position the FMN ligand from the UbiX crystal structure within the pocket of the HmfG homology model. [101] The two structures are overlaid in Figure 18.

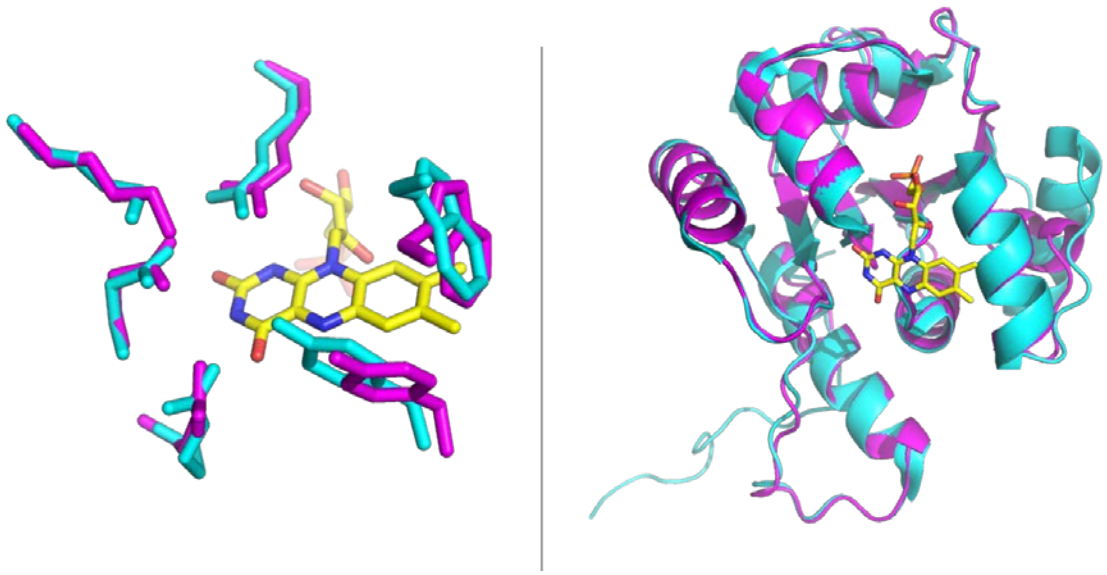


Figure 18: Structural alignment of UbiX and the HmfG homology model. UbiX crystal structure; HmfG homology model. Key active site residues from UbiX are shown on the left with their counterparts from the HmfG homology model: (clockwise from center top) R122, W200, Y169, R139, E140, and K129. S15, E49, S90, and R185 are also conserved, but not shown. The alignment between protein monomers is shown on the right.

Many of the active site residues in UbiX have been assigned proposed functions. [91] These residues are entirely conserved between UbiX and HmfG. They are listed with their proposed function in Table 6. As Figure 18 illustrates, the conserved residues also occupy similar positions within the UbiX crystal structure and the HmfG homology model. Overall, HmfG was found to have 40.86 % sequence identity with a sequence similarity of 0.39 and a coverage of 0.79, based on the HHblits algorithm. [100]

Table 6: Residues that are critical to UbiX function are conserved in HmfG.

Residue in UbiX	Proposed Function	Conservation in HmfG
S15	Acid-base catalysis on FMN N5	Identical → S52
E49	Acid-base catalysis on FMN N5	Identical → E87
S90	DMAP phosphate binding	Identical → S114
R122	DMAP phosphate binding	Identical → R146
K129	DMAP phosphate binding	Identical → K153
R139	DMAP phosphate binding	Identical → R163
E140	DMAP phosphate binding	Identical → E164
Y169	Carbocation stabilization	Identical → Y193
R185	DMAP phosphate binding	Identical → R209
W200	Carbocation stabilization	Identical → W225

3.2.5 Discussion

The homology structures predicted for HmfF and HmfG were based on relatively low sequence identities with their structural templates 2IDB and 1SBZ, respectively. Nevertheless, the resulting structures overlaid well enough with Fdc1 and UbiX to place key residues determined to be critical to the function of each enzyme in the same locations. The phenotypical similarities between the HmfF and HmfG knockouts explored by Koopman, et al. (2010) [52] and the UbiD and UbiX knockouts that helped to define the family in the 1970s and 1980s suggest similar functionality between the enzymes. [88-89] The complete conservation of functional residues, along with the high degree of predicted structural alignment suggests that these functional similarities may be the result identical catalytic mechanisms. This evidence supports the hypotheses that (1) HmfF is a UbiD-like prFMN-dependent reversible decarboxylase, and (2) HmfF is a UbiX-like flavin prenyltransferase.

3.3 Enzyme Expression, Purification, Catalysis, and Characterization

3.3.1 Statement of Authorship

The text in this section was authored by Harrison Bellow Rose. All experiments and data analysis were performed by the author, with assistance from undergraduate research assistant Aimee C. Moise [a] and research scientist John M. Robbins. [a]

[a] School of Chemical & Biomolecular Engineering Georgia Institute of Technology, Atlanta, Georgia 30332, United States.

3.3.2 Abstract

The enzymes listed in Table 3 were expressed in *Escherichia coli*. All enzymes other than HMFO and UbiX were evaluated for catalytic activity in either whole-cell, cell lysate, or purified protein. 2,3-DHBD_Ao, 2,6-DHBD_Rs, and SAD_Tm were found to have no activity on FDCA. The reported activities of HmfF and HmfG were verified in whole-cell catalysis and lysate, and preliminary results suggested this activity to be reversible. Neither HmfF nor HmfG could be purified in an active form. HMFO was expressed and purified but not evaluated for catalytic activity.

3.3.3 Introduction

The literature review in section 3.1 Biocatalyst Candidate Selection provided a list of seven potential catalysts for the conversion of furfural to FDCA (summarized in Table 3). Three of these, 2,3-DHBD_Ao, 2,6-DHBD_Rs, and SAD_Tm are cofactor-free

reversible aromatic decarboxylases that have been characterized previously, but have not been tested for activity on furans. [75] One enzyme, HmfF, is a putative FDCA decarboxylase that is hypothesized to require a prenylated flavin cofactor. Two enzymes, HmfG and UbiX, are proposed to prenylate FMN, producing the cofactor for HmfF. HmfG has not been characterized, but UbiX has been studied in detail. [91, 95] At the start of this study, however, the existence of prFMN was not known, and there was no postulated mechanism for the HmfF/HmfG system. The final enzyme in the list, HMFO, is an engineered HMF oxidase variant that has been demonstrated to have high catalytic activity towards FFA, and to be inactive on furfural. [71]

The experiments that follow were designed to validate these enzymes for use as biocatalysts. Preliminary screening was performed using whole-cells or cell lysate. Enzymes that failed the screening were discarded, and those that passed were developed further for purification and free-enzyme biocatalysis. UbiX and HMFO have been validated for their required activities as described elsewhere, [71, 91] and were not subject to screening in this work.

3.3.4 Methods

3.3.4.1 Cloning of Enzymes for Heterologous Expression

Codon-optimized genes for 2,3-DHBD_Ao, 2,6-DHBD_Rs, and SAD_Tm in pET-21a expression vectors were provided by Kurt Faber and Silvia Glueck of the Faber research group at the Austrian Centre of Industrial Biotechnology, the Department of Chemistry, Organic & Bioorganic Chemistry, University of Graz, Graz, Austria. [75] The

expression construct for 2,6-DHBD_Rs contained an N-terminal histidine tag. A pNIC28a-Bsa4 vector containing the gene for UbiX with an amino-terminal histidine tag and TEV protease cleavage site was provided by Karl Payne and David Leys of the Leys research group at the Centre for Synthetic Biology of Fine and Specialty Chemicals, Manchester Institute of Biotechnology, The University of Manchester, Manchester, UK.[91] The expression systems for for 2,3-DHBD_Ao, 2,6-DHBD_Rs, SAD_Tm, and UbiX were used as provided, and no modifications were made. Codon-optimized genes for HmfF, HmfG, and the HMFO double-variant engineered by Dijkman, et al. (2015)[71] were purchased from Genscript (Picastaway, NJ) as listed in the appendix section A.2 Protein DNA Sequences. Each gene was purchased with leading and lagging restriction sites and were delivered in pUC57 vectors.

HmfF, HmfG, and HMFO were ligated into a variety of expression plasmids for a series of expression tests with and without solubility or purification tags. A hexahistidine repeat was used as a nickel affinity purification tag, and a small ubiquitin-like modifier (SUMO) peptide was used as a solubility enhancing tag.[102] When tags were used in this work, they were placed on the amino terminus of each enzyme. Site-directed mutagenesis was used to change restriction sites where necessary, and overlap extension polymerase chain reaction (PCR) was used to construct genes for fusion proteins (i.e., {hexahistidine tag}-{SUMO tag}-{enzyme}). Details of the PCR reactions performed in this work are provided in Table 7. Primer sequences are provided in the appendix section A.3 Primer Sequences. Endonucleases corresponding to the restriction sites listed were used to cleave sequences for insertion into vectors.

Table 7: Details of PCR performed in this work. Primers sequences are listed in the appendix section A.3 Primer Sequences.

PCR Template(s)	Primer IDs	Resultant Product
{E}-{HmfF}-{H}-<pUC57>	1, 3	{D}-{HmfF}-{H}
{E}-{HmfF}-{H}-<pUC57>	2, 3	{N}-{†HmfF}-{H}
{D}-{HmfF}-{H}-<pET-27b> {N}-{6xHis SUMO}-<pET-28b>	A, B, 4, 5	{N}-{6xHis SUMO}-{‡HmfF}-{H}
{D}-{HmfG}-{X}-<pET-17b> {N}-{6xHis SUMO}-<pET-28b>	A, B, 6, 7	{N}-{6xHis SUMO}-{‡HmfG}-{X}
{D}-{HMFO}-{X}-<pET-28a> {N}-{6xHis SUMO}-<pET-28b>	A, B, 8, 9	{N}-{6xHis SUMO}-{‡HMFO}-{X}

E = EcoRI restriction site: (G▾AATTC)
H = HindIII restriction site: (A▾AGCTT)
D = NdeI restriction site: (CA▾TATG)
N = NcoI restriction site: (C▾CATGG)
X = XhoI restriction site: (C▾TCGAG)
{xyz} = Restriction site, gene, or T7 terminator or promoter
<xyz> = Vector
† = Glycine codon `GCC` inserted after start codon
‡ = Start codon removed

Site-directed mutagenesis was used to revise the leading EcoRI site on HmfF to an NdeI site for restriction and subsequent ligation into the multiple cloning site (MCS) of pET-27b for untagged expression. An additional round of site-directed mutagenesis was used to revise the leading restriction site again from NdeI to NcoI for restriction and ligation into the leading MCS of pETDuet for untagged expression in tandem with HmfG. HmfG was restricted and ligated directly into the MCS of pET-17b and pET-21c vectors, and into the second MCS of pETDuet. HMFO was restricted and then ligated directly into pET-28a as an intermediate step before overlap-extension PCR.

Overlap-extension PCR was used to construct the genes for fusion proteins of each enzyme with leading affinity and solubility tags. An in-house pET-28b plasmid with the affinity and solubility tag sequence (provided in the appendix section A.2.4 Hexahistidine Affinity Tag and SUMO Solubility Tag) in the MCS was used as the tag template. The pET-27b, pET-17b, and pET-28a plasmids containing HmfF, HmfG, and HMFO, respectively were used as the templates for each enzyme gene. The fusion sequences prepared in this manner were restricted at the leading NcoI restriction site and the lagging HindIII or XhoI site as applicable and ligated into pET-28b for expression.

Overall, eleven expression constructs were used or prepared in this work, including those that were provided by other laboratories or prepared as described above. The full list of expression constructs is provided in Table 8. A plasmid map of construct 10, SUMO-tagged HmfG in the MCS of a pET-28b vector is provided in Figure 19.

Table 8: Expression constructs used or prepared in this work.

No.	Enzyme(s)	Tag	Vector	Source/Preparation
1	2,3-DHBD_Ao	Untagged	pET-21a	Faber Research Group
2	2,6-DHBD_Rs	6xHis	pET-21a	Faber Research Group
3	SAD_Tm	Untagged	pET-21a	Faber Research Group
4	UbiX	6xHis	pNIC28a-Bsa4	Leys Research Group
5	HmfF	Untagged	pET-27b	Prepared as described
6	HmfG	Untagged	pET-17b	Prepared as described
7	HmfG	Untagged	pET-21c	Prepared as described
8	HmfF & HmfG	Untagged	pETDuet	Prepared as described
9	HmfF	6xHis SUMO	pET-28b	Prepared as described
10	HmfG	6xHis SUMO	pET-28b	Prepared as described
11	HMFO	6xHis SUMO	pET-28b	Prepared as described

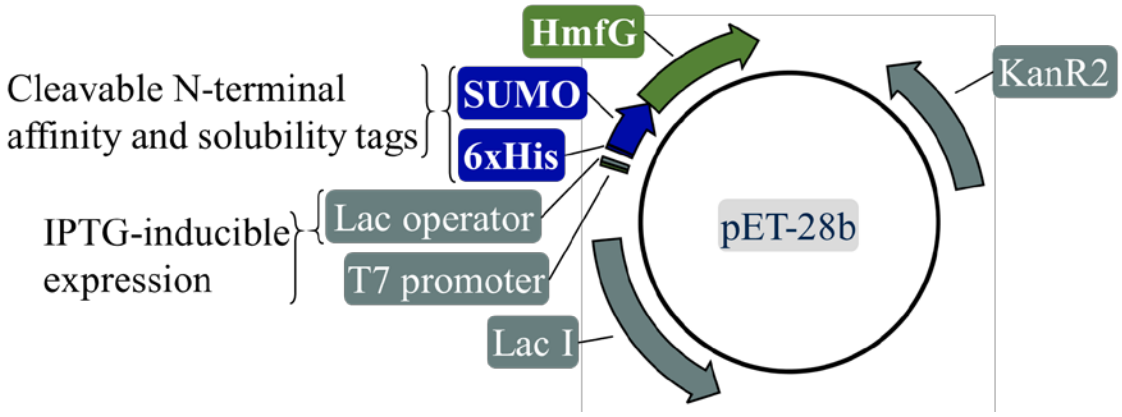


Figure 19: Plasmid map of affinity- and solubility-tagged HmfG. This construct corresponds to entry 10 in Table 8, and consists of a pET-28b with a hexahistidine-SUMO-HmfG fusion ligated into the MCS.

3.3.4.2 Heterologous Protein Expression

Each of the expression constructs in Table 8 was transformed into chemically competent *Escherichia coli* BL21 DE3 or Rosetta2 DE3 for expression. In some experiments, both constructs 5 and 7 were co-transformed as an alternate to construct 8 for dual expression. Transformed bacteria were plated onto 1.5 % agar with 25 g L^{-1} Luria-Bertani broth (LB) and either ampicillin (50 µg mL^{-1}) or kanamycin (30 µg mL^{-1}), depending on the vector. Colonies were used to inoculate 5 mL cultures of 25 mg L^{-1} LB and antibiotic as before. Both ampicillin and kanamycin were used for dual-expression in which both constructs 5 and 7 were co-transformed. Chloramphenicol was also used for expression in Rosetta2 DE3. Expression cultures were either 5 mL in glass culture tubes, or 50 mL, 100 mL, 500 mL, or 1 L in culture flasks. Expression cultures were grown to an optical density (at 600 nm) of between 0.6 and 0.7, and induced with 40 µM isopropyl β-D-1-thiogalactopyranoside (IPTG). Expression was conducted under varied combinations

of time and temperature. After expression, cells were harvested by centrifugation in up to six 250 mL bottles for 15 minutes at 3,836 g and 4 °C. Cell pellets were decanted, refilled with remaining culture, and centrifuged again until all culture had been processed, eventually yielding up to six large cell pellets. Decanted pellets were optionally lyophilized.

In experiments requiring cell lysis, such as for evaluation of expressed protein, pelleted cells were frozen at -80 °C for at least 45 minutes, or for up to two days before use. Subsequent purification steps were performed in two-pellet batches. Two-pellet batches were resuspended in 50 mL lysis buffer (sodium phosphate (50 mM), sodium chloride (300 mM), imidazole (10 mM, optional), pH 8.0) with 0.1 mg mL^{-1} of hen egg white lysozyme (HEW, EC# 235-747-3, Amresco LLC, Solon, OH) and cells were lysed by ultrasonic homogenization in an ice-water bath. Cell lysate was distributed into two 35 mL conical tubes and clarified by centrifugation at 23,200 g and 4 °C for 20 minutes. The resulting supernatant and sediment were taken to be the “soluble” or “clarified” and “insoluble” lysate fractions, respectively. Samples from each fraction were then evaluated by denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

3.3.4.3 Whole-cell and Lysate Biocatalysis

2,3-DHBD_Ao, 2,6-DHBD_Rs, and SAD_Tm were expressed as described by Wuensch, et al. (2012) [75] and evaluated for whole-cell catalytic activity on FA (decarboxylation) and furfural (carboxylation) using the established protocol. [74, 76] The model substrates for all three enzymes: orcinol (for carboxylation) and

2,6-dihydroxymethylbenzoic acid were used as positive controls. *E. coli* lacking exogenous plasmid was used as a negative control. Reactions were allowed to proceed for 24 hours at 30°C in sealed microcentrifuge tubes. Reactions were quenched by a rapid freeze-thaw cycle in liquid nitrogen and cell mass was removed by centrifugation. Samples were acidified using trifluoroacetic acid (TFA) and conversion was evaluated by high-performance liquid chromatography (HPLC) using the established method on a C18 column. [76]

HmfF and HmfG were evaluated for catalytic activity using a similar procedure. Constructs 5 and 7 were expressed in *E. coli* BL21 DE3 as described, for 5.5 hours at 25 °C after induction with 0.04 mM IPTG. Both constructs were expressed individually, as well as together in a dual-plasmid system. Lysate and whole cells from individual expressions were also mixed after expression as an alternative to dual-expression. Whole and lysed cells without exogenous plasmids were used as a negative control, in addition to negative controls devoid of any biological material. Both decarboxylation of FDCA (the reported native activity) and carboxylation of FA (the reverse reaction) were evaluated. The full combinatorial study design is illustrated in Figure 20. Carboxylation of furfural and decarboxylation of FFA were also evaluated.

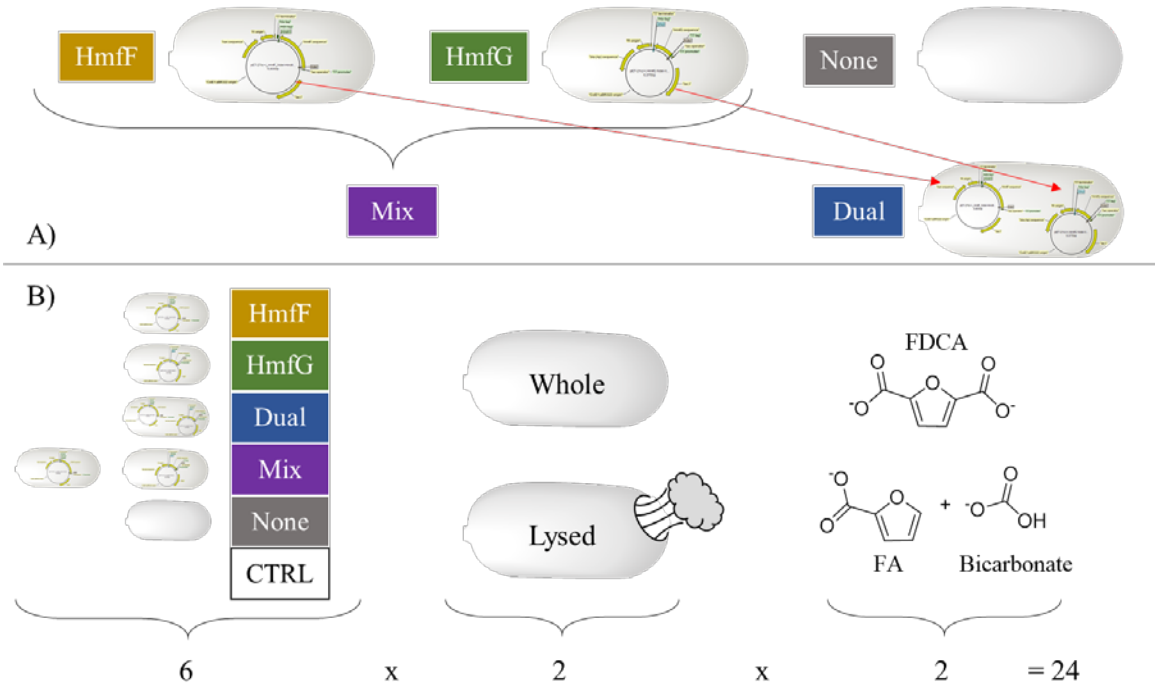


Figure 20: Combinatorial design of experiment for activity of HmfF and HmfG. (Panel A): *E. coli* were transformed with plasmids encoding either HmfF (, "HmfF") or HmfG (, "HmfG") for individual expression, or with both plasmids for dual expression (, "Dual"). Cells were also combined after individual expression (, "Mix"). Untransformed cells were used as a negative expression control (, "None"). (Panel B): An additional control lacking any *E. coli* was also used ("CTRL"). Cultures were either kept as whole-cells or lysed by ultrasonic homogenation, and either FDCA was added to evaluate the native decarboxylation activity, or FA and bicarbonate were added to evaluate the reverse reaction. Overall, 24 conditions were evaluated, in addition to controls without substrate.

Organic substrates were provided at 10 mM, and inorganic carbon was provided as 50 mM sodium bicarbonate for the carboxylation experiments. Reactions were carried out in 100 mM potassium phosphate buffer, pH 7.5 in sealed microcentrifuge tubes, which were incubated at 37 °C for 16 hours. Reactions were quenched with a rapid and repeated freeze-thaw cycle with liquid nitrogen.

Samples were evaluated by HPLC. 1 % TFA was added to each sample, and 10 μL injections were made onto a C18 column followed by isocratic elution with 89.9 % water, 10 % acetonitrile, and 0.1 % TFA. Relative conversion was determined by chromatogram peak area corresponding to product formation for either the decarboxylation or carboxylation reactions. Furfural, FA, FFA, and FDCA were evaluated at 276 nm, 244 nm, 287 nm, and 263 nm, respectively.

3.3.4.4 Affinity Chromatography and Protein Purification

Histidine-tagged proteins (translation products of constructs 2, 4, and 9-11) were purified for individual characterization. Clarified lysate was prepared from expression cultures as described previously. Insoluble cell debris was discarded, and clarified lysate was pooled and incubated on ice for 30 minutes with 2-4 mL of Novagen Ni-NTA His·BIND RESIN (EMD Millipore Corporation; Merck KGaA, Darmstadt, Germany). After incubation, resin was collected on a polyethylene column, or by centrifugation at 35 g and 4 °C for 5 minutes and subsequently washed twice: first by resuspension in 10 mL of lysis buffer (sodium phosphate (50 mM), sodium chloride (300 mM), imidazole (10 mM), pH 8.0), and then in 10 mL of wash buffer (sodium phosphate (50 mM), sodium chloride (300 mM), imidazole (20 mM), pH 8.0). The washed resin was resuspended a final time in 2.5 mL of elution buffer (sodium phosphate (50 mM), sodium chloride (300 mM), imidazole (500 mM), pH 8.0) to displace the bound protein, and collected once more. Purified protein was decanted by pipette from centrifuged resin or collected as column effluent. Protein was then optionally passed through a 0.2 μm syringe filter to remove

misfolded aggregates, and also optionally dialyzed or processed for buffer exchange on a PD10 size exclusion chromatography column.

In investigations requiring cleavage of the histidine and SUMO tags (such as in constructs 9-11), purified proteins were exchanged into standard buffer (sodium phosphate (50 mM), sodium chloride (300 mM), pH 8.0). SUMO tags were cleaved at using an in-house histidine-tagged SUMO-specific protease (ULP1) during a 1-hour incubation at 4 °C. Uncleaved protein, cleaved tags, and his-tagged protease were removed by binding onto 2 mL of fresh Ni-NTA resin as described previously. The resin was pelleted or collected on a clean column, and the supernatant or effluent containing the cleaved protein was collected and optionally passed through a 0.2 µm syringe filter to remove misfolded aggregates. A simplified cartoon schematic of the full purification scheme is provided in Figure 21.

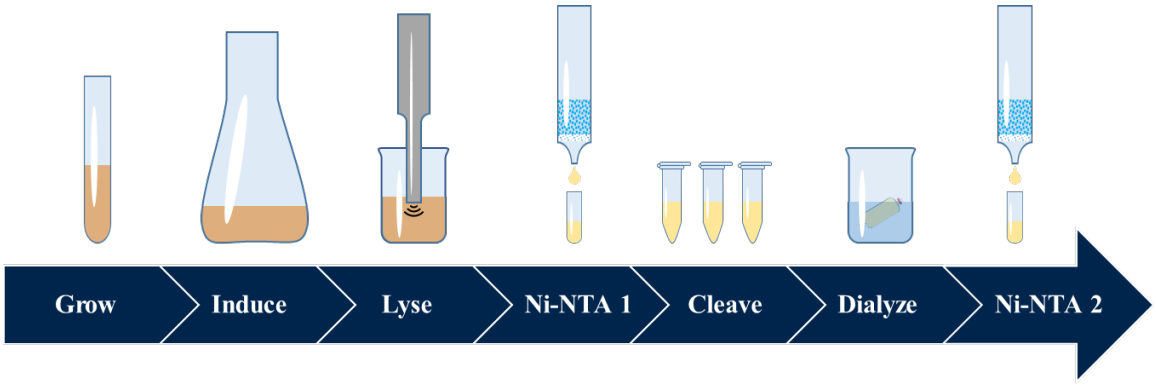


Figure 21: Full expression and purification protocol for His-SUMO- tagged enzymes. Transformed *Escherichia coli* are passaged up to culture flasks and protein expression is induced with IPTG. After expression, cells are lysed by ultrasonic homogenization. Clarified lysate is incubated with nickel and washed. Affinity and solubility tags are optionally cleaved from purified protein and buffer exchange is performed to remove imidazole. A second chromatographic step removes separates the cleaved enzyme from uncleaved enzyme, cleaved tags, and protease.

3.3.5 Results

3.3.5.1 Heterologous Protein Expression

Constructs 1-3 were expressed as described by Wuensch, et al. (2012) [75] and soluble expression was achieved for all three enzymes 2,3-DHBD_Ao, 2,6-DHBD_Rs, and SAD_Tm. UbiX was expressed using Construct 4 as described by White, et al. (2015) [91] and soluble expression was achieved.

Soluble expression of HmfF was achieved using construct 5 in BL21, with optimal expression at 25 °C. Total expression of HmfF decreased slightly in the dual-plasmid system with constructs 5 and 7. Individual expression was improved with the addition of the SUMO tag (construct 9), and optimal expression of the tagged construct was achieved at 18 °C.

No expression of HmfG could be detected using construct 6, and only insoluble expression was achieved using construct 7, with greater expression levels observed at 30 °C than at 18 °C, and the highest levels observed at 25 °C. Overall expression decreased in the dual-plasmid system with constructs 5 and 7 – some insoluble expression was observed at 25 °C, but no expression was observed after expression at 18 °C. Soluble HmfG was only achieved with the addition of the SUMO tag (construct 10), and optimal soluble expression was achieved at 25 °C.

Overall, no substantive improvements in the expression of HmfF or HmfG were observed in the dual-plasmid expression system. This result is inconsistent with the hypothesis that HmfF is a folding chaperone for HmfG, but consistent with functions suggested by the Fdc1/UbiX homology (i.e., which suggest no need for direct physical contact between the two proteins). The SUMO-fusions, however, yielded improved soluble expression for both enzymes. Because of these results, the untagged tandem expression system using pETDuet (construct 8) was not evaluated.

Soluble expression of SUMO-tagged HMFO was achieved with construct 11 using the methods described previously. Better soluble expression was achieved at 18 °C than at 25 °C or 30 °C. This finding of better expression at lower temperatures is consistent with the protocol reported by Dijkman and Fraaije (2014) [58] which involved a 68 hour expression at 17 °C in Terrific Broth.

3.3.5.2 Whole-cell and Lysate Biocatalysis

The reported reversible 2,6-dihydroxybenzoic acid decarboxylation activities of 2,3-DHBD_Ao, 2,6-DHBD_Rs, and SAD_Tm in whole-cell catalysis were confirmed by HPLC using the published methods. [75-76] No carboxylation activity was observed using furfural as a substrate, nor was any activity observed on FA. This result is consistent with the lack of published activities on heterocycles. [74-75] This result is also consistent with the proposed decarboxylation mechanism, which is initiated by proton abstraction of a phenolic hydroxyl group adjacent to the carboxylate group. [76]

Individually expressed HmfF was found to decarboxylate FDCA to FA in whole-cell and in lysate, although relative conversion was 3-4 times greater in the case of whole cells than with lysate (based on two independent trials). An example set of chromatograms from one experimental trial is provided in Figure 22. No appreciable decarboxylation activity was observed with HmfG alone, either in whole-cells or lysate. Whole cells with dual expression of both HmfF and HmfG together ranged from 90 % to 230 % as effective as those with overexpression of HmfF alone, but a mixture of whole cells from cultures overexpressing HmfF and HmfG individually consistently overperformed whole cells with HmfF alone, although the improvement varied from one and a half to three-fold. These results are consistent with the observations made by Koopman, et al. (2010) [52] that HmfG is inactive on its own, but enhances the decarboxylation activity of HmfF in cell extract. These results also support the hypothesis that HmfG is a flavin prenyltransferase and HmfF is a UbiX-like prFMN-dependent decarboxylase, although the discovery of prFMN had not yet been published at that time. The baseline decarboxylation activity of individually expressed HmfF may be indicative of prFMN production by an endogenous flavin prenyltransferase within *E. coli*. Interestingly, the mixture of both single-plasmid cultures consistently showed the greatest conversion, suggesting that – if HmfG is in fact a flavin prenyltransferase – unnecessary prFMN may be discarded into the reaction media by the HmfG-producing cells, and taken up by the HmfF-producing cells. Horizontal gene transfer of plasmids between cultures is another possibility, and one that should be investigated if this system is adapted for two-strain whole-cell biomanufacturing. This second hypothesis

could reasonably be investigated by repeating the same experiment and then plating the mixture of both strains onto agar prepared with both ampicillin and kanamycin.

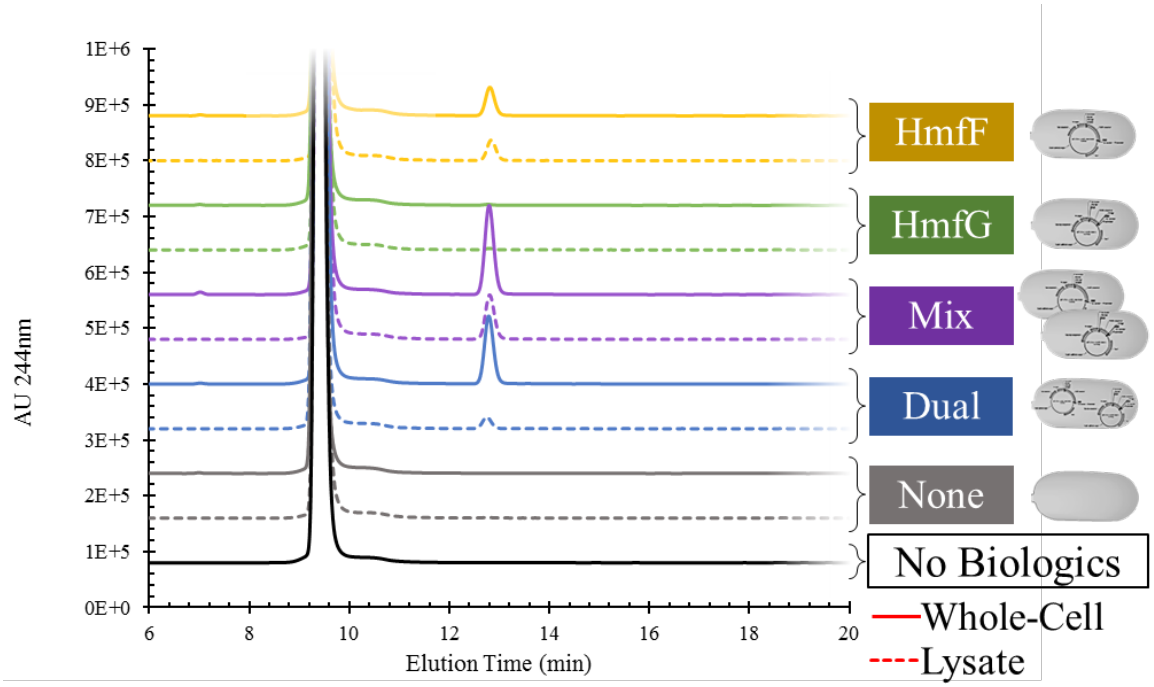


Figure 22: Stacked HPLC chromatograms tracking the decarboxylation of FDCA by HmfF and HmfG. Product (FA) formation is evaluated by the area of absorbance peaks at 244 nm and 12.8 minutes. Unreacted substrate (FDCA) is visible at 9.4 minutes. Series labels correspond to the different experimental conditions described in Figure 20. All reactions were started together and quenched together, so relative conversion can be determined by comparison of the FA peak areas which are proportional to product concentration.

In all cases, cell lysate underperformed whole-cells. It is not possible to determine the cause of this decrease in conversion based on the available data, but some reasonable hypotheses can be posed: (A) it may be the case that the total amount of active enzyme was lower in the lysate experiments due to protein damage that occurred during ultrasonic

homogenization. (B) complete formation and maturation of the prFMN^{iminium} cofactor requires a series of reduction and oxidation steps to occur in a particular sequence. [92] Therefore, if HmfF is in fact a prFMN-dependent decarboxylase, the decrease in catalytic activity of the lysate may be the result of disruption of the cellular environment making successful production and maturation of the prFMNiminium less likely.

Finally, there is no evidence from these data to suggest that movement of substrates and products across the cell membrane is a limiting factor in substrate conversion. Active and passive transport mechanisms were not determined or evaluated.

HmfF was found to carboxylate FA to FDCA as well, supporting the hypothesis that the decarboxylation reactions in this family are reversible, as proposed in Figure 7 and discussed in the literature. [72-76, 94-95] In fact, FDCA was formed spontaneously without catalysis at about 30-50 % of the rate demonstrated by whole-cell mixture of individually expressed HmfF and HmfG. Lysed or whole cells expressing HmfG alone provided no additional conversion relative to this base case devoid of any biological material, which is consistent with the lack of decarboxylation activity previously described. The overall trend of a reduction in catalytic activity after cell lysis that was observed for decarboxylation of FDCA was also observed for carboxylation. No activity was observed in the absence of bicarbonate. These results are summarized in Figure 23.

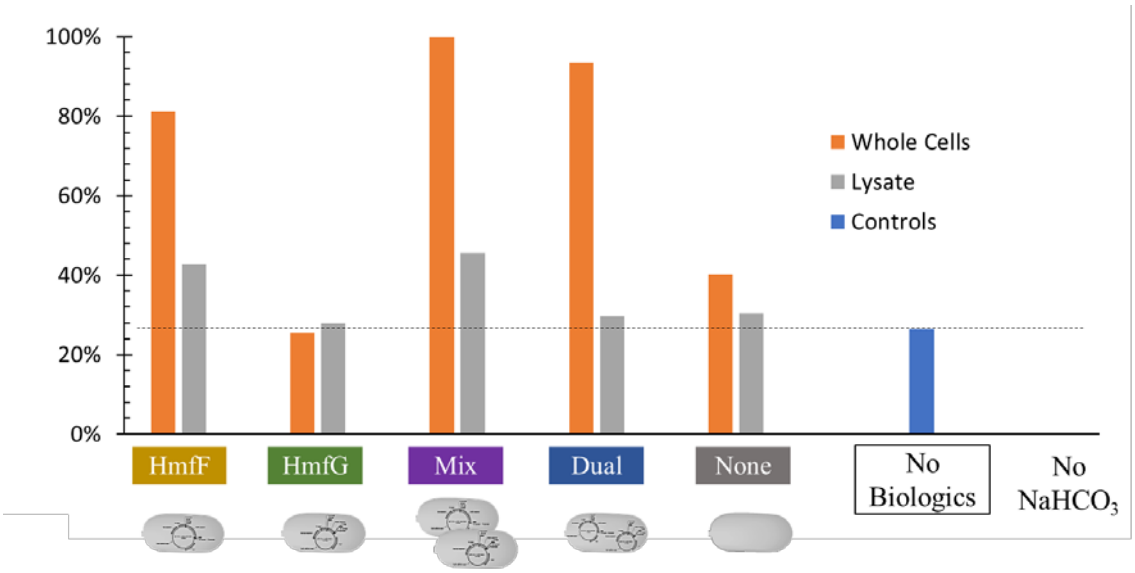


Figure 23: Normalized conversion of FA to FDCA by HmfF and HmfG in whole cells and lysate. Relative product (FDCA) formation was determined by HPLC as described. Series labels correspond to the different experimental conditions described in Figure 20. A control without any biological material shows that some level of carboxylation may occur spontaneously under these conditions. No FDCA is formed in the absence of bicarbonate.

Catalysis experiments evaluating carboxylation of furfural to FFA (the desired activity necessary for the catalytic route from furfural to FDCA discussed in Figure 8) were inconclusive due to difficulties in chromatographic analysis.

3.3.5.3 Affinity Chromatography and Protein Purification

Histidine-tagged 2,6-DHBD_Rs was purified from lyophilized cell culture as described (although eluted into 250 mM imidazole, rather than 500 mM) and found to be soluble and stable. The established whole-cell catalytic activites of 2,6-DBD_Rs on orcinol and 2,6-dihydroxy-4-methylbenzoate were verified with the purified free-enzyme directly in elution buffer. No catalytic activity was observed on furfural or FFA, which is consistent

with the results of the whole-cell catalysis experiments and mechanism discussed previously.

SUMO-tagged HmfF (construct 9) was purified generally without complication. It was found that cleavage of the SUMO tag could be performed concomitantly with buffer exchange by dialysis when buffers were supplemented with 10 % glycerol.

Tagged HmfG (construct 10) was found to be recalcitrant and difficult to purify. The tagged enzyme was found to aggregate during affinity chromatography, making the use of filter columns impracticable due to clogging, even when purifications were conducted in a cold room at 4 °C. The high concentration of imidazole in the elution buffer provided an apparent stabilizing effect; buffer exchange by dialysis or size exclusion chromatography also caused near complete aggregation. This aggregation could be reduced by replacing the imidazole with equimolar arginine during dialysis, but arginine was found to prevent cleavage of the SUMO tag by SUMO-specific protease. In addition to enabling soluble expression as noted previously, the SUMO tag also improved protein stability during purification, and tag cleavage resulted in aggregation of HmfG.

Several additional stabilization strategies were attempted in several combinations including: expression in alternate strains of *Escherichia coli* such as Rosetta2 DE3, supplementation of growth media with riboflavin, purification into buffers with and without arginine or glycerol, and cleavage while column-bound, or in the presence of stabilizing agents such as Brij L23 (Croda interational PLC, Millipore Sigma).

Ultimately, soluble cleaved HmfG was produced only by outpacing the kinetics of aggregation with streamlined rapid purification (cell lysis to product in less than 10 hours) and periodic removal of aggregates by gentle syringe filtration. Solutions of purified protein, however, were observed to become turbid within an hour.

UbiX (construct 4) was purified according to the protocol established by White, et al. (2015) [91] although cells were lysed as described above (i.e., with a combination of freeze-thaw cycle, HEW, and ultrasonic homogenization) rather than with a French Press cell disruptor. UbiX was found to exhibit similar aggregation behavior to HmfG, although aggregation was resolved with membrane filtration and fast purification – two steps that were apparently necessary but insufficient for the stabilization of HmfG.

Tagged HMFO (construct 11) was successfully purified using the same protocol that was described previously. All purifications of HMFO were performed in a cold room at 4 °C. Some aggregation was observed when simultaneous tag cleavage and dialysis were performed as was described for HmfG, so that procedure is not recommended for HMFO. Alternate purification protocols for HMFO are available in the literature. [71]

3.3.5.4 Enzyme Characterization

In addition to SDS-PAGE, purified proteins were characterized by other methods. SUMO-tagged HmfG was consistently yellow when it was eluted after affinity chromatography (Figure 24). The yellow component could not be isolated for further study.

Figure 24: Effluent from affinity chromatography shows that tagged HmfG is yellow. Concentrated protein elutes as a yellow plug from a nickel affinity column. Subsequent elution fractions of histidine-tagged SUMO-HmfG fusion are shown from left to right. The middle sample has the highest protein concentration.

Several attempts were made to concentrate the yellow protein and isolate the yellow component, but they were ultimately unsuccessful. Tagged HmfG concentrated on a centrifugal membrane filter became colorless, and no colored species were found in the filtrate. Tagged yellow protein purified into buffer without glycerol was concentrated by lyophilization or by SpeedVac, but in both cases, protein aggregated and the apparent color was lost. Absorbance spectra could not be obtained from soluble protein due to the rapid aggregation, and characteristic flavin peaks were not observed in boiled protein extract.

The apparent tendency of HmfG to purify with a yellow component is consistent with the presence of a bound colored cofactor or prosthetic group within the enzyme. Inductively coupled plasma optical emission spectrometry (ICP-OES) conducted by Katherine Vest in the Cobine Laboratory (Department of Biological Sciences, Auburn University, Auburn Alabama) was negative for bound metals or metal clusters in either

HmfF or HmfG, suggesting that the colored constituent is likely a conjugated organic structure. Samples of HmfG that were supplemented with FMN were found to produce a yellow pellet after aggregation. This result is consistent with the homology studies and suggests that HmfG does in fact strongly bind flavin. Combined, the metals analysis and FMN pull-down assay support the hypothesis that the colored species that copurifies with HmfG is in fact a flavin or related cofactor.

The tricyclic isoalloxazine head of the FMN molecule (and the tetracyclic prFMN) are heavily conjugated and are known to absorb light at different wavelengths depending on the oxidation state. Oxidized FMN is yellow, neutral and anionic semiquinone states are blue and red, respectively, and colorless when reduced. [103] In contrast, prFMN that is reoxidized is reportedly purple. [93] It is unlikely that dissociated FMN would become reduced and colorless in the aerobic laboratory environment, as reducing agents like sodium dithionite are typically required to produce $FMNH_2$ outside of cells or anaerobic chambers.

The loss of color after membrane concentration would suggest cofactor dissociation, except that is inconsistent with both the FMN pull-down results and the fact that HmfG elutes with the yellow component in the first place. Additionally, no color was observed in the filtrate or in boiled extract, and dissociated flavin would likely oxidize to a colored state as mentioned above. Collectively, these results suggest that the loss of color is not due to dissociation of cofactor. An alternative possibility that is supported by the data is the occurrence of some chemical transformation within the enzyme resulting in the loss

of color. This is a plausible explanation, as HmfG is already believed to catalyze chemical reactions of flavins.

Untagged samples of purified HmfF and HmfG, and purified UbiX were evaluated in triplicate by differential scanning fluorimetry (DSF) on a NanoTemper Prometheus NT.48 (NanoTemper Technologies, Inc. South San Francisco, California). HmfF and UbiX were suspended in standard buffer solution, and HmfG was in standard buffer supplemented with 500 mM arginine. The ratio of fluorescence at 350 nm/330 nm was recorded over a temperature scan from 20 °C to 95 °C with a temperature ramp rate of 1 °C/minute. Results are reported in Table 9, and were found to be sensitive to concentration, although that apparent relationship was not probed in more detail. HmfG was found to have a similar melting point to dilute UbiX, and both were found to exhibit two inflections in fluorescence ratio. These findings are consistent with the hypothesis that UbiX and HmfG have homologous structures, as discussed in section 3.2.4.2, HmfG Structural Prediction and Alignment with UbiX.

Table 9: Melting points of HmfF, HmfG, and UbiX by DSF. Melting points were taken to be the inflection point of the fluorescence ratios at 350 nm/330 nm. An apparent concentration dependence was observed for the melting transition of UbiX.

Protein Construct	$T_{m,1}$ (°C)	$T_{m,2}$ (°C)
Cleaved HmfF	45.57 ± 0.03	NA
Cleaved HmfG	44.5 ± 0.1	77.6; ID
Tagged UbiX, dilute	46.6 ± 0.6	78.93 ± 0.09
Tagged UbiX, concentrated	64.1 ± 0.08	NA
NA = not applicable. ID = insufficient data for uncertainty estimation.		

3.4 Discussion

HmfF and HmfG are homologues to a model system (i.e., UbiX and Fdc1) that produce and use a modified flavin cofactor, prFMN. When this work began, prFMN had not yet been discovered, and it is likely the key to many more examples of interesting chemistry that have yet to be described. The data suggest that prFMN is also the key to the unique and industrially important reversible FDCA decarboxylase activity that has been demonstrated for HmfF. If HmfF can be developed with further study, it may serve as the first and enduring example and application of this new cofactor in industrial processes. Further study of this family of enzymes may lead to other promising targets for direct carboxylation and production of renewable products that were previously limited by the lack of necessary chemistry.

CHAPTER 4. THERMODYNAMIC CHARACTERIZATION OF FURAN-SERIES COMPOUNDS

4.1 Introduction

The preliminary thermodynamic analysis performed for the two alternative reaction pathways (i.e., oxidation and then carboxylation: $\mathrm{Fur} \xrightarrow{8} \mathrm{FA} \xrightarrow{9} \mathrm{FDCA}$, or carboxylation and then oxidation: $\mathrm{Fur} \xrightarrow{17} \mathrm{FFA} \xrightarrow{15} \mathrm{FDCA}$) presented in Figure 7 revealed that aqueous carboxylation is likely to be equilibrium-limited. This was an expected result only insofar as it did not suggest that carboxylation is *highly favored*; even the possibility of reversibility was counterintuitive, given the high oxidation state and low energy of CO_2. [72] The uncertainty of those estimates, however, underscores the need to develop a better understanding of the thermodynamic limitations of the reaction system and to identify those conditions which are likely to provide the greatest thermodynamic drive towards carboxylation. In other words, more accurate data and more precise mathematical tools are necessary to predict what set of reaction conditions for carboxylation provides the highest equilibrium constant and lowest reaction quotient. These conditions – whatever they may be – define the space within which the search for furfural carboxylation is most likely to be successful.

The defining criterion for thermodynamic equilibrium is that the Gibbs energy of the system be minimized. For a single reaction occurring in a closed single-phase system, this point is represented as the thermodynamic reaction equilibrium constant, K_{eq}^{th}, which

is the product of activities of participating species, a_i, raised to their stoichiometric coefficients, ν_i, as shown in Equation (3). Thermodynamic activity is an effective concentration, equal to the product of concentration, C_i, and activity coefficient, γ_i. [104] Equation (3) thus illustrates that any purported relationship between equilibrium constants and concentration alone is an approximation based on the implicit assumption that all activity coefficients are equal to unity (i.e., an idealized system).

$$K_{eq}^{th} = \prod_i a_i^{\nu_i} = \prod_i (C_i \gamma_i)^{\nu_i} \tag{3}$$

Mathematical representations of chemical reaction equilibria become more complicated in the context of solution chemistry, even in the seemingly straightforward cases of electrolytes dissolved into water. Bioreactors typically fall into this category, and it is worth noting that any enzymatic conversion of furfural, FA, FFA, or FDCA is likely to occur in aqueous solution with additional buffer species and even dissolved gasses present. In aqueous buffer systems, additional equilibrium relationships must be established for the ionic dissociation reactions of each electrolyte species, and a charge balance must be included in addition to the general species balance. However, the degree of dissociation of each ionizable species is itself a function of ionic strength. Mathematical representations of aqueous ionization and neutralization reactions are discussed in more detail in CHAPTER 5.

Additional complexity is engendered when multiple phases are present, such as when gases are sparged into a reaction vessel or solid precipitates are formed or consumed.

The influence of these conditions on reactions are not always obvious, and mathematical tools are frequently necessary for their analysis. The selection of reaction conditions in which multiple phases are present can often make or break a chemical process. Precipitation events can enable continuous reactive crystallization, which may permit complete conversion of what would otherwise be an unfavorable reaction. [80, 105] The accidental formation of solid or gaseous phases in process equipment not designed for handling such flows, however could have disastrous results. [106]

While these relationships may be predicted if the heats and free energies of formation for all solubilized species in a system are known, such information is not available for many organic acids and bases, and reference states used in calculation are sometimes left ambiguous in the literature. Solution heats and energies of formation may be determined from standard heats and energies of formation, if the heat capacity and heats and energies of solvation are known. These properties in turn may be determined by calorimetry, or estimated using computerized force field-based simulations such as density functional theory. [70, 107-116] Locating values for thermodynamic properties can be very time consuming, so multiple databases have been established to curate thermodynamic data including GAMPHI, the Minnesota Solvation Database, and the Dortmund Data Bank. [117-119]

To make thermodynamic predictions for systems that do not conform to standard state conditions for which tabulated data are available, then an equation of state is needed. Kenneth Pitzer extended previous methods for predicting activity coefficients of aqueous electrolyte systems by incorporating Debye-Hückel theory to evaluate forces between pairs

of solvated ions. [120-131] Many group contribution models exist, such as that used to make the aqueous heat of formation predictions in MetaCyc that are listed in Table 2, but different methods specialize in predicting properties of different classes of compounds. Generally, their applicability to organic compounds solvated in aqueous solution is limited. [132] Hybrid modeling approaches for more complex aqueous mixtures, including dicarboxylic acids, have been explored by the aerosol and atmospheric sciences community. [133-134]

In 2000, Gross and Sadowski adapted the Statistical Associating Fluid Theory (SAFT), an equation of state which represents molecules as jointed chains of hard spheres to incorporate perturbations accounting for the difference in dispersal contributions of bonded and non-bonded spheres. [135] Later, universal model constants for n-alkanes were introduced, and the model was subsequently dubbed Perturbed-Chain SAFT (PC-SAFT). [136] In a collaboration with members of the Sadowski research group, we have used PC-SAFT to capture the activities of furfural, FA, FFA, and FDCA. This work is described in the following section: 4.2, Mutual Influence of Furfural and Furancarboxylic Acids on Their Solubility in Aqueous Solutions: Experiments and PC-SAFT Predictions.

4.2 Mutual Influence of Furfural and Furancarboxylic Acids on Their Solubility in Aqueous Solutions: Experiments and PC-SAFT Predictions

4.2.1 Statement of Authorship

The text in this section was coauthored by the following: Harrison Bellow Rose, [a] Thorsten Greinert, [b] Christoph Held, [b] Gabriele Sadowski, [b] and Andreas S. Bommarius. [a, c]

[a] School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States. [b] Laboratory of Thermodynamics, Department of Biochemical and Chemical Engineering, TU Dortmund University, 44227 Dortmund, Germany. [c] School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States.

Unless otherwise referenced, all experimental data presented were collected by Harrison B. Rose, with assistance from undergraduate research assistant Madison M. Wilber. All PC-SAFT analysis and plotting of data was performed by Thorsten Greinert and Christoph Held.

Harrison B. Rose and Thorsten Greinert contributed equally to this work.

4.2.2 Abstract

Furan compounds are of mounting global interest due to their biorenewable nature and their potential to replace petroleum-based compounds as feedstocks in manufacturing. In this work the solubilities of furfural and the furancarboxylic acids 2-furoic acid (FA), 5-formyl-2-furancarboxylic acid (FFA), and 2,5-furandicarboxylic acid (FDCA) in aqueous solutions and organic solvents were investigated experimentally and by modeling with perturbed-chain statistical associating fluid theory (PC-SAFT). The PC-SAFT pure-component parameters of the solutes FA, FFA, and FDCA and one binary parameter between each solute and each solvent were adjusted to fit experimentally determined solubilities of each solute in each organic solvent or in water. Pure-component parameters of furfural were fitted to experimental density data and vapor-pressure data, and a binary

interaction parameter was fitted to capture the solubility behavior of furfural in water. Modeling of pH effects enhanced predictions of the mutual influences of the acids on their solubilities in ternary aqueous systems. Mutual solubility influences of furfural and the furancarboxylic acids were accurately modeled with one constant binary parameter for the acid-furfural mixtures. All PC-SAFT modeling results were validated with new experimental solubility data at 35 °C, which were measured by HPLC analysis of equilibrated saturated solutions.

4.2.3 Introduction

Furan compounds such as furfural and many furancarboxylic acids are of mounting global interest due to their biorenewable nature and their potential for substitution into industries and processes traditionally built upon petroleum-based compounds.[4] Enzymatic valorization and whole-cell biotransformation of these furans are enticing process targets due to their promise of mild operating conditions, but the chemical and thermodynamic information necessary for process development is often still not available. An improved understanding of the chemical and thermodynamic behaviors of these species is a prerequisite both for efficient adaptation of existing processes, and for development of new processes that involve these building blocks.

In particular, standard data and activity coefficients are necessary to estimate the limiting conversion (and thus the feasibility) of desired reactions, [137] prior to investing in the development of biocatalysts. When details of solubility and phase equilibria are known,

high product yields can sometimes be coaxed even from thermodynamically limited biotransformations, through product partitioning or precipitation-driven synthesis. [80, 105]

Some thermodynamic data are available on the constituents of the furan series, such as the phase diagram of furfural in water, [138-140] the heats of formation and combustion of furfural and 2-furoic acid (FA), [113, 115-116, 141] and the melting point, pKa, solubility, and enthalpy of solution of FA in water and some binary solvent mixtures. [107, 112, 142-143] Gaussian and ADF predictions of the free energy of solvation of furfural and 2,5-furandicarboxylic acid (FDCA) are also available, [108] as well as the solubility of FDCA in water-acetic acid mixtures. [144]

This work aims to supplement the available information on furfural and the furancarboxylic acids that may serve as a basis for future efforts in aqueous-phase reaction engineering of mixtures containing furan series compounds. To that end, we present the solubilities of furfural, FA, FDCA, and 5-formyl-2-furancarboxylic acid (FFA) in binary and ternary aqueous solutions through experimentation and modeling.

Ternary mixtures involving aqueous organic solutes have been successfully described using methods such as UNIFAC, [145-146] NRTL, [147] hard-sphere model, [147] and SAFT EOS. [148] The Perturbed-Chain SAFT (PC-SAFT) [136, 149] has been demonstrated to be effective in capturing the solution thermodynamics of aqueous sugar systems. [150] For that reason, PC-SAFT was selected to predict solubility in aqueous solutions containing either ethanol or other organic cosolvents, or furancarboxylic acids as cosolutes. As the mixtures under investigation are further prone to pH-effects, a combined approach for

mutual-interaction effects assisted by pH influences was proposed in this work following the procedure developed by Daldrup, et al. (2011) [151] for aqueous amino-acid solutions.

4.2.4 Methods

4.2.4.1 Chemicals

Furfural, minimum 98 % FA (Sigma-Aldrich, St. Louis, MO), FFA, and FDCA (TCI America, Portland, OR) were used as the furan series compounds in this study. 200 proof ethanol (DLI, King of Prussia, PA) was used for solvent preparation. Optima LC/MS-grade water and methanol (both from Fisher Chemical, Fair Lawn, NJ) were used for solvent and sample preparation and analytical chromatography. ACS-grade potassium phosphate (VWR BDH Chemicals, Westchester, PA), and 85.0 % phosphoric acid (EMD Millipore Corporation, Billerica, MA) were also used for chromatography. 10N sodium hydroxide (Fisher Chemical) was used for sample pH adjustment where stated. Manufacturers' stated purity for ethanol, furfural, FA, FFA, FDCA, and water, the compounds evaluated in this study are listed in Table 10. No additional purification or analyses of purity were performed on purchased materials prior to use.

Table 10: Chemical Sample Table

Chemical Name	Source	Initial Mole Fraction Purity
Ethanol	DLI	≥ 0.999
Furfural	Sigma-Aldrich	≥ 0.99
FA	TCI America	≥ 0.98
FFA	TCI America	≥ 0.980
FDCA	TCI America	≥ 0.980
Water	Fisher Chemical	$\gg 0.999$

4.2.4.2 Sample Preparation and Handling

Single-solute samples of either FFA or FDCA were prepared in triplicate by their addition to 1-2 mL of water, ethanol, or to 25 %, 50 %, or 75 % (by weight) mixtures of water and ethanol in 2 mL microcentrifuge tubes. Additional samples of FA, FFA, or FDCA in water were titrated by the addition of sodium hydroxide to achieve saturated solutions over a range of pH values. Pure-solvent samples without pH adjustment were prepared at 18, 25, 32, 37, and 45 °C, while mixed-solute and pH-adjusted samples were prepared and evaluated at 35 °C. In all cases, the solutes were added beyond their saturation limits, resulting in the visible accumulation of insoluble material. The microcentrifuge tubes were closed to prevent evaporation and samples were incubated at the required temperature for 48 hours with continuous vigorous mixing (1.3-1.4 kRPM). After 48 hours, samples were centrifuged at 14 kRPM and visually examined for presence of an insoluble pellet. If no insoluble material was observed, more solute was added and the sample was incubated for another 48 hours. Samples with a visible pellet were decanted by pipette into a syringe with a 0.2 μm filter and filtered into a new microcentrifuge tube which was returned to the samples' incubation temperature, but without mixing. Pipette tips, syringes, and syringe filters, and microcentrifuge tubes for filtrate were all pre-heated to the sample temperature, if it was greater than room temperature. Filtered saturated solutions were serially diluted into the initial solvent until they were within the linear range of detection by optical spectroscopy. All dilutions were performed on an analytical balance to ensure accurate accounting by weight and to avoid calculation errors due to density

changes. Solubility data (and pH values of saturated solute+water solutions) for pure- and mixed-solute samples without pH adjustment are provided in Table 21, in the appendix section B.1, Single-Solute Data. Solubility data and corresponding pH values for samples with pH-adjustment are provided in Table 22, in the appendix section B.1, Single-Solute Data.

Two-solute samples were handled in a similar manner, although their initial preparation and ultimate analysis differed. Saturated solutions of FA, FFA, or FDCA in water were prepared and filtered at 35 °C. Aliquots of the filtered solutions were diluted to 33 % and 66 % by weight. Triplicate samples were then prepared combinatorically using these 33 %, 66 %, or 100 % solutions as the solvent. That is, three aqueous samples each containing an excess amount of one of the studied solutes (either a furancarboxylic acid or furfural) in a solvent of known cosolute (either one carboxylic acid or furfural) concentration were prepared. All samples were incubated at 35 °C for 3.5-5 days, then inspected, filtered, and diluted as described previously. The saturation concentrations (the solubility) of the solutes in these samples were evaluated by reverse-phase high-performance liquid chromatography (HPLC). The weight fraction of the saturated solutes at the respective cosolute weight fractions are listed in Table 23, in the appendix section B.2, Mixed-Solute Data.

Additional ternary samples were prepared in a similar manner for pH measurement. In more detail, three aqueous samples each containing excess amount of one of the studied solutes (either a furancarboxylic acid or furfural) in a solvent of known cosolute (either one carboxylic acid or, furfural) concentration were prepared. Although the equilibrium

pH values of these additional samples were determined experimentally, their final compositions were not measured directly. Instead, the modeling dataset was completed based on the ternary samples in Table 23 that were evaluated by HPLC. As samples were prepared under controlled conditions, samples' compositions were related to those HPLC samples that were prepared under identical or bounding conditions. For example: the saturating FFA weight fraction of a sample prepared using a 100 % solution of furfural was taken to be that of the HPLC sample prepared under the same conditions. The saturating FDCA weight fraction of a sample prepared using a 50 % solution of FA was linearly interpolated from the saturating FDCA weight fractions of the HPLC samples prepared using 33 % and 66 % solutions of FA. pH measurements of the ternary samples, and the cosolute weight fractions are listed in Table 24, in the appendix section B.2, Mixed-Solute Data.

4.2.4.3 UV-Visible Spectroscopy

Optical absorbance measurements of single-analyte samples were performed on a DU 800 Spectrophotometer (Beckman Coulter, Inc.). The instrument was blanked with the bulk solvent of the sample of interest (i.e., everything but the analyte), and baseline correction was performed. Samples containing FFA were analyzed at 280 nm and those containing FDCA were analyzed at 260 nm. Separate standard curves were prepared for FFA and FDCA in each solvent composition. All measurements were performed in a clean quartz cuvette with a path length of 1 cm. Standard solutions were prepared by dilution of 100 mM stocks to establish extinction coefficients and the linear range of detection. As

samples and standards were diluted in two serial steps into the linear range, sample concentrations were back-calculated from their measured absorbance.

4.2.4.4 HPLC

Multi-analyte samples were evaluated by reverse-phase HPLC (Shimadzu Scientific Instruments) using a DGU-20A5 membrane degasser, a LC-20AT solvent delivery unit, a SIL-20AC HT autosampler, a CTO-20A column oven, a SPD-M20A photodiode array detector, and a CBM-20A system controller (all components, Shimadzu Scientific Instruments). A Hyperclone 5u MOS (C8) 120A 125x4.6 cm column (Phenomenex, Inc.) was used for component separation. The mobile phase consisted of 90 % (v/v) 20 mM potassium phosphate, pH 2.0, and 10 % (v/v) methanol and was delivered at 1 mL per minute. Standards were prepared by diluting filtered saturated stock solutions for each analyte into water. Each standard was prepared from stock in triplicate. Samples and standards were diluted 1/200 into mobile phase prior to loading into the autosampler, and 1 μL of each standard or sample was injected onto the column. Sample composition was determined by linear deconvolution of the resultant peaks.

4.2.4.5 pH Measurements

Ag/AgCl pH electrode (Thermofisher Scientific) with an Accumet Research AR50 pH meter (Fisher Scientific). Electrode standardization and sample measurements were conducted at 35 °C.

4.2.4.6 PC-SAFT Modeling of Solubilities

Based on the phase equilibrium conditions for solid and liquid phases – assuming pure solid phases and neglecting the temperature dependence of the heat capacities – the solid solubility of component i at atmospheric pressure was calculated according to Prausnitz (1969) [152] and Gmehling, et al. (1978) [153]:

$$x_i^L = \frac{1}{\gamma_i^L} exp\left[-\frac{\Delta H_{fus,i}}{RT}\left(1 - \frac{T}{T_{fus,i}}\right)\right] \quad (4)$$

The quantities $\Delta H_{fus,i}$ and $T_{fus,i}$ represent the enthalpy and the temperature of melting of the pure substance i, respectively. The influence of the medium on the solubility of component i is described only by the activity coefficients (γ_i^L), which are dependent on composition and temperature. Original PC-SAFT was used to model these activity coefficients. In original PC-SAFT, the residual Helmholtz energy is calculated as the sum of different contributions to the residual Helmholtz energy $A^{residual}$ such as hard-chain repulsion $A^{hard\ chain}$, dispersion (van-der-Waals) $A^{dispersion}$, and association $A^{association}$ caused by hydrogen bonding.

$$A^{residual} = A^{hard\ chain} + A^{dispersion} + A^{association} \quad (5)$$

The equations for hard-chain and dispersion contributions have been reported previously.[136, 149] The association contribution was used as suggested.[149] To describe an associating compound, five pure-component parameters are required: the segment number

(m_i^{seg}), the segment diameter (σ_i), the dispersion-energy parameter (u_i/k_B) with the Boltzmann constant k_B, the association-energy parameter ($\epsilon^{A_iB_i}/k_B$), and the association-volume parameter ($\kappa^{A_iB_i}$). In addition, the number of association sites N^{assoc} (association scheme) must be set prior to calculations. The latter was not treated as a fitting parameter.

To describe binary systems, the conventional Berthelot-Lorentz combining rules were applied. Additionally, one binary parameter k_{ij} was introduced correcting the dispersion-energy parameter for the mixture of components i and j according to Equation (7):

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{6}$$

$$u_{ij} = (1 - k_{ij})\sqrt{u_i u_j} \tag{7}$$

The binary interaction parameter k_{ij} corrects for deviations of the cross-dispersion energy u_{ij} from the geometric mean of the self-dispersion energy parameters u_i and u_j. The cross-association energy between two associating substances i and j can be described by applying mixing and combining rules as suggested by Wolbach and Sandler (1998) [154]:

$$\epsilon^{A_iB_j} = \frac{1}{2}(\epsilon^{A_iB_i} + \epsilon^{A_jB_j}) \tag{8}$$

$$\kappa^{A_iB_j} = \sqrt{\kappa^{A_iB_i}\kappa^{A_jB_j}}\left(\frac{\sqrt{\sigma_i\sigma_j}}{0.5(\sigma_i + \sigma_j)}\right)^3 \tag{9}$$

The availability of $A^{residual}$ from Equation (5) finally allows calculating thermodynamic properties (e.g., fugacities) as described elsewhere. [136]

To account for pH effects on the solid solubility of the furancarboxylic acids in water, an approach was used which has previously been demonstrated for amino acids in aqueous solutions. [155] To this end, the deprotonation of the furancarboxylic acids can be characterized by the following acid dissociation constants:

$$K_{a1} = \frac{a(A^-)a(H_3O^+)}{a(HA^\pm)} \tag{10}$$

$$K_{a2} = \frac{a(A^{2-})a(H_3O^+)}{a(A^-)} \tag{11}$$

The acid dissociation constant K_{a1} is calculated from the activity of the non-dissociated acid $a(HA^\pm)$ and the activities of the dissociated species which are the acid $a(A^-)$ and the hydrogen ion which binds to a water molecule $a(H_3O^+)$. The acid dissociation constant K_{a2} is calculated analogously including the activity of the acid after two dissociation steps $a(A^{2-})$. The acid constants K_a were used in Equation (12) to calculate the solid solubility, $x^L_{HA,total}$, from the solid solubility of the non-dissociated acid, $x^L_{HA,\pm}$, and the solution pH. Equation (12) is valid at low pH (in this case, up to pH ≈ 3.5).

$$x^L_{HA,total} = x^L_{HA,\pm}\left(1 + \frac{K_{a,1}}{10^{-pH}} + \frac{K_{a1}K_{a2}}{(10^{-pH})^2}\right) \tag{12}$$

The solid solubility of the non-dissociated acid was calculated with PC-SAFT using Equation (4).

The liquid solubility of furfural in water was calculated based on Equation (13), which is the conventional liquid-liquid equilibrium condition for two liquid phases L' and L''.

$$x_i^{L'}\gamma_i^{L'} = x_i^{L''}\gamma_i^{L''} \tag{13}$$

4.2.5 Results

4.2.5.1 Solubility Data, pH, and pK_a

Using the materials and methods described, a detailed dataset was collected to supplement the standard thermodynamic data already available for the furan series compounds, including experimental data for solubility, pH, andpK_a. The individual solubilities of FFA and FDCA were measured without pH adjustment in water and in ethanol from 18-45 °C, as well as in 1:3, 1:1, and 3:1 mixtures (by weight) of ethanol and water at 35 °C. The individual solubilities of FA, FFA, and FDCA were also at 35 °C in water with pH adjustment. The solubilities and equilibrium pH values of ternary systems consisting of any two of the four furan series compounds in water were also measured at 35 °C.

Concomitant to this study, pKa values were measured for FFA and FDCA at 25 °C and 0.1 M ionic strength (see CHAPTER 5, section 5.3.6.2, Spectrophotometric

Determination of pK_a Values). These values were extrapolated to infinite dilution using a general modified Debye-Hückel relationship. [156] The thermodynamic pK_a value for FFA was found to be 2.22, while the first and second pK_a values of FDCA were found to be 2.06 and 3.44, respectively. A value of 3.16 (infinite dilution, 35 °C) was used for FA. [112]

4.2.5.2 PC-SAFT Parameter Estimation

Prior to parameter estimation the association schemes for the furan compounds were determined. The carboxylic group was characterized by one hydrogen-donor site and one hydrogen-acceptor site. The cyclic O-group within FA and FFA as well as the formyl-group within FFA were assigned by one additional hydrogen-acceptor site each.

The PC-SAFT parameters for FA, FFA, and FDCA were fitted to experimental solid solubilities in organic solvents. Five pure-component parameters and one k_{ij} value between these solutes and the considered organic solvent were adjusted by minimizing the objective function OF1 using a number, NP, of experimental data points.

$$OF1 = \frac{1}{NP}\sum_{k=1}^{NP}\left(1 - \frac{x_i^{L,PC-SAFT}}{x_i^{L,exp}}\right)_k^2 \tag{14}$$

Note that the temperature range for the parameter estimation was 280-326 K for FA and 298-318 K for FFA and FDCA, respectively. This is a typical temperature range for parameterization of biomolecules' PC-SAFT parameters (e.g., [157-151]); nevertheless, the temperature extrapolation to higher temperatures has not been validated in this work.

PC-SAFT modeled solid solubilities $x_i^{L,PC-SAFT}$ were calculated according to Equation (4) using melting properties from the literature with references given in Table 11. Experimental solid solubilities $x_i^{L,exp}$ of solutes in organic solvents were measured in this work or taken from literature with the references from Table 12.

The PC-SAFT parameters for furfural were fitted to literature-density data and vapor-pressure data of pure furfural. The three PC-SAFT parameters were adjusted based on the objective function OF2:

$$OF2 = \frac{1}{NP}\sum_{k=1}^{NP}\left(1-\frac{\rho_k^{PC-SAFT}}{\rho_k^{exp}}\right)^2+\left(1-\frac{p_k^{PC-SAFT}}{p_k^{exp}}\right)^2 \quad (15)$$

Experimental densities ρ_k^{exp} and vapor pressures p_k^{exp} were taken from the literature. [158-159]

Table 11: Pure-component and binary PC-SAFT parameters for furancarboxylic acids, furfural, and solvents.

Solute	m_i^{seg}	σ_i (Å)	u_i/k_B (K)	N_i^{assoc}
FA [this work]	4.3304	2.3929	133.52	3B (2+1)
FFA [this work]	8.7974	3.1607	149.86	4B (3+1)
FDCA [this work]	2.3743	3.2393	342.15	4C (2+2)
Furfural [this work]	3.0711	3.3558	320.07	-
Water [160]	1.2047	[c]	353.95	2B (1+1)
Ethanol [149]	2.3827	3.1771	198.24	2B (1+1)
Acetone [161]	2.8913	3.2279	247.42	2B (1+1)
Ethyl ethanoate [136]	3.5375	3.3079	230.80	2B (1+1)
Propan-1-ol [149]	2.9997	3.2522	233.40	2B (1+1)

Solute	$\epsilon^{A_iB_i}/k_B$ (K)	$\kappa^{A_iB_i}$	$T_{fus,i}$ (K)	$\Delta H_{fus,i}$ (kJ mol^{-1})	k_{ij} (H_2O)
FA [this work]	500.00	0.1000	402.5 [115]	22.6 [115]	-5. 59E-02
FFA [this work]	2107.92	0.0001	473.2 [162]	26.7[a]	-2.06E-01
FDCA [this work]	228.25	0.0001	615.2[b]	34.1[d]	-4.00E-03
Furfural [this work]	-	-	234.5[163]	14.4[164]	-8.17E-03
Water [160]	2425.67	0.0451	-	-	-
Ethanol [149]	2653.40	0.0324	-	-	-3.10E-02 [165-166]
Acetone [161]	-	0.0003	-	-	-
Ethyl ethanoate [136]	-	0.0100	-	-	-
Propan-1-ol [149]	2276.80	0.0153	-	-	-

[a] Obtained from Scifinder, data were obtained from Syracuse Research Corporation of Syracuse, New York (US).
[b] Calculated with Walden Rule ($\Delta H_{fus} = \Delta S_{fus} \cdot T_{fus}$) using $\Delta S_{fus} = 56.5$ J/(mol K) which is an established value. [167]
[c] For water, a temperature-dependent segment diameter $\sigma = 2.7927 + 10.11 \times \exp(-0.01775T) - 1.417 \times \exp(-0.01146T)$ was used.
[d] Estimated with Joback estimation method taken from Dortmund Data Bank (http://ddbonline.ddbst.com/OnlinePropertyEstimation/OnlinePropertyEstimation.exe)

The obtained parameters are listed in Table 11. It is known that some of the PC-SAFT parameters are correlated (e.g., m^{seg} and σ, ε^{AiBi} and κ^{AiBi}). The compounds

considered in this work do not present a sufficiently high number for a parameter-correlative analysis. Thus, the PC-SAFT parameters of the considered compounds were treated as freely adjustable parameters without any constraints fulfilling OF1 or OF2.

The average absolute deviation (AAD, see Equation (16)) and the average relative deviation (ARD, see Equation (17)) between PC-SAFT modeled and experimental solid solubility data are given in Table 12.

$$AAD = \frac{1}{NP}\sum_{k=1}^{NP}\left|\left(y_k^{PC-SAFT} - y_k^{exp}\right)\right| \tag{16}$$

$$ARD = \frac{1}{NP}\sum_{k=1}^{NP}\left|\left(1 - \frac{y_k^{PC-SAFT}}{y_k^{exp}}\right)\right| \cdot 100\% \tag{17}$$

Table 12: Binary PC-SAFT parameters between furancarboxylic acids and solvents. Deviations between correlated and experimental solid solubilities in binary systems of furancarboxylic acids and solvents are also given.

Solute	Solvent	T-range (K)	ARD (%)	AAD (%)	k_{ij} $(solvent)$
FA	Ethanol [143]	279-325	0.91	0.0039	-6.30E-02 [a]
FA	Acetone [143]	280-316	1.79	0.0057	-3.27E-02 [a]
FA	Ethyl ethanoate [143]	281-322	2.44	0.0041	1.86E-02 [a]
FA	Propan-1-ol [143]	283-326	1.53	0.0050	-2.02E-02 [a]
FA	Furfural [this work]	-	-	-	-2.80E-01 [b]
FA	Water [this work]	-	-	-	-5.59E-02 [a]
FFA	Ethanol [this work]	298-318	1.40	0.0010	-4.55E-02 [a]
FFA	Furfural [this work]	-	-	-	-2.80E-01 [b]
FFA	Water [this work]	-	-	-	-2.06E-01 [a]
FDCA	Ethanol [this work]	298-318	7.70	0.0008	-1.19E-01 [a]
FDCA	Furfural [this work]	-	-	-	-2.80E-01 [b]
FDCA	Water [this work]	-	-	-	-4.00E-03 [b]

[a] Obtained from fit to binary systems.
[b] Obtained from fit to ternary systems.

Figure 25 compares liquid-density data from Liessmann, et al. (1995) [158] and from Daubert and Danner (1985) [159] with PC-SAFT. The densities decrease with increasing temperature which is accurately described with PC-SAFT. This is quantified by the ARD value of 0.6 %.

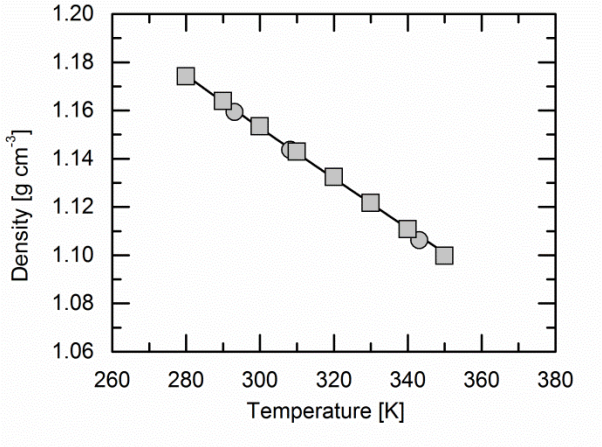


Figure 25: Atmospheric liquid-density of furfural vs. temperature. Symbols are experimental data (○, Liessmann, et al. (1995) [158]; □, Daubert and Danner (1985) [159]), solid line represents density modeling with PC-SAFT using the parameters of Table 11.

Figure 26 illustrates that the vapor pressures of furfural were modeled successfully with PC-SAFT compared to experimental values from Daubert and Danner,[159] which is quantified by the ARD value of 0.5 %.

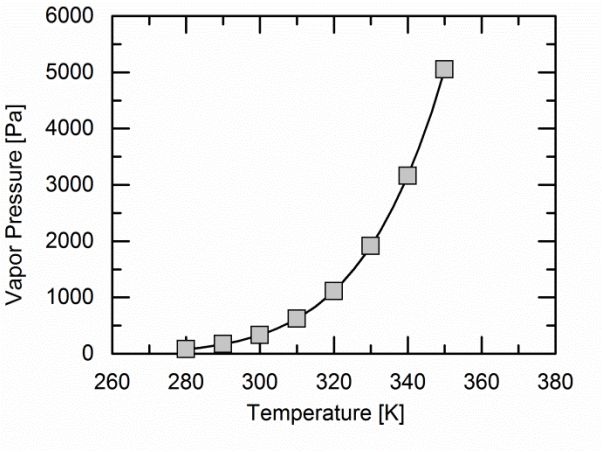


Figure 26: Vapor pressure of furfural vs. temperature. Symbols are experimental data from Daubert and Danner (1985) [159], solid line represents vapor pressure modeling with PC-SAFT using the parameters in Table 11.

In the following sections, all so-determined PC-SAFT parameters are validated by modeling and predicting solubility of binary and ternary systems.

4.2.5.3 Solid Solubility in Non-Aqueous Solvent

Figure 27 compares the experimentally measured solid solubilities of solute/ethanol mixtures for FA, FFA, and FDCA with PC-SAFT correlation results. Experimental data for FA are from Guo, et al. (2014) [143] while those for FFA and FDCA are from this work. PC-SAFT pure-component parameters as well as one binary interaction parameter between each solute and ethanol were adjusted to fit the experimental data. As expected, the solid solubility of each of the three solutes increases with increasing temperature. The absolute solid solubilities as well as the temperature influence on the solid solubilities increase in the sequence FDCA<FFA<FA. Unlike dissolution into water where acid dissociation tends

to be more appreciable, ethanol provides an environment that suppresses acid dissociation and makes trends in intrinsic solubility and stability of the pure solid phase more accessible to experimental inquiry. The solubility trend in Figure 27 consistently reflects an inversion of the trends of molar mass, T_{fus}, and ΔH_{fus}, all of which increase in the order FA<FFA<FDCA and are general indicators of the relative stabilities of each pure acid. The solid solubilities and the temperature dependence are described accurately with PC-SAFT, resulting in ARD values of 0.9 % for FA, 1.4 % for FFA, and 7.7 % for FDCA, respectively. Please note that the ARD for FDCA is comparatively large, which is a mathematical consequence from the very small absolute solubility values of FDCA in ethanol (one order of magnitude lower compared to FA or FFA). This can be further seen on the AAD value for FDCA in ethanol, which is the smallest one for the systems considered in Figure 27.

Parameter fitting for FA also incorporated published data for its solid solubility in acetone, ethyl ethanoate, and propan-1-ol. These results are not shown here graphically, as the data was only used to adjust the PC-SAFT parameters. The deviations between these correlations and the experimental data are listed as AAD and ARD values in Table 12, along with the data references.

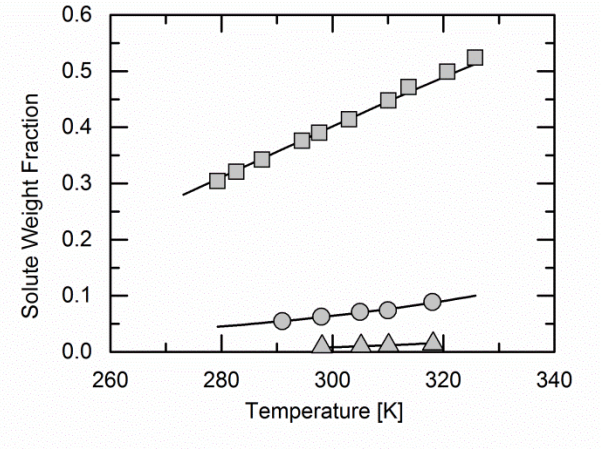


Figure 27: Solid solubility of FA, FFA, and FDCA in ethanol vs. temperature at 1 bar. Symbols are experimental data from this work (○, FFA; △, FDCA) and from literature (□, FA) with references as listed in Table 12, and the solid lines represent PC-SAFT correlations using the parameters in Table 11 and Table 12.

4.2.5.4 Individual Solubilities in Water and pH Dependence

The liquid solubility of furfural in water at 35 °C was modeled using PC-SAFT by fitting the binary interaction parameter (Table 11) to the furfural solubility data from Table 21.

The solid solubilities of the furancarboxylic acids in water at 35 °C and a range of pH are shown in Figure 28. An additional k_{ij} parameter between each acid and water was estimated based on the solid solubility of the acid in the non-dissociated state in the binary mixture furancarboxylic acid + water. This value was converted from the absolute cosolute-free solubility $x^L_{HA,total}$ as listed in Table 22 to $x^L_{HA,\pm}$ using the pH values from Table 22 and Equation (12), which describes the dependence of the solid solubility of an acid in

water on pH. These intrinsic solubilities, which are the low horizontal asymptotes for each series in Figure 28, follow the same overall trend that was observed for the non-dissociated acids in ethanol (Figure 27). As the pH of each binary solution is increased, the total acid solubility increases. In each case, this increase is attributed to the deprotonation of the solubilized neutral acid into its much more soluble charged form, allowing more neutral acid to become solubilized, up to the intrinsic limit $x^{L}_{HA,\pm}$ and satisfying the constraint the acidity constant places on the ratio of the deprotonated acid to its neutral form. This increase in net solubility ($x^{L}_{HA,total}$) continues until a stable equilibrium between pH and acid concentration is reached.

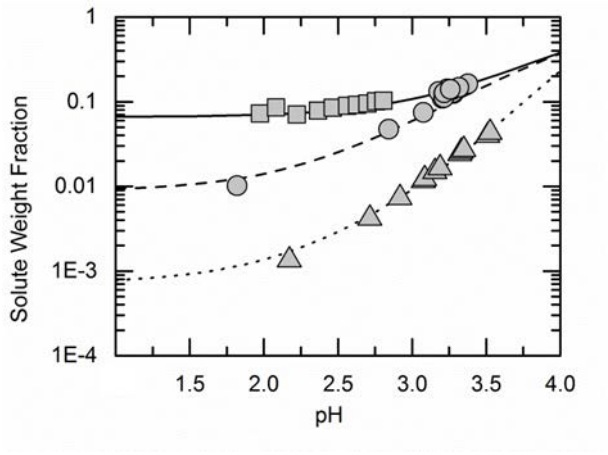


Figure 28: Solid solubility of FA, FFA, and FDCA in water at 35 °C and 1 bar as functions of pH. Symbols are experimental data from this work (□, FA; ○, FFA; △, FDCA), lines (solid line, FA; dashed line, FFA; dotted line, FDCA) represent calculations based on Equation (12) using the solid solubility of the non-dissociated acid $x^{L}_{HA,\pm}$.

4.2.5.5 Solid Solubility in Water/Organic Cosolvent Mixtures

Figure 29 shows the experimental dependence of the solid solubility of FDCA in ethanol/water mixtures on the FDCA-free solvent composition. The solid solubility at $w_w^{solute-free} = 1.0$ is not the experimental value but the solid solubility of the non-dissociated FDCA, which was calculated with Equation (12) from the experimental value $x_{HA,total}^L$. For experimental solid solubility values in solvent with $w_w^{solute-free} < 1.0$ it was assumed that no dissociation takes place and therefore the experimental values are solid solubilities of the uncharged species. The PC-SAFT predicted solubility values are, accordingly, also solubility values of the non-dissociated FDCA.

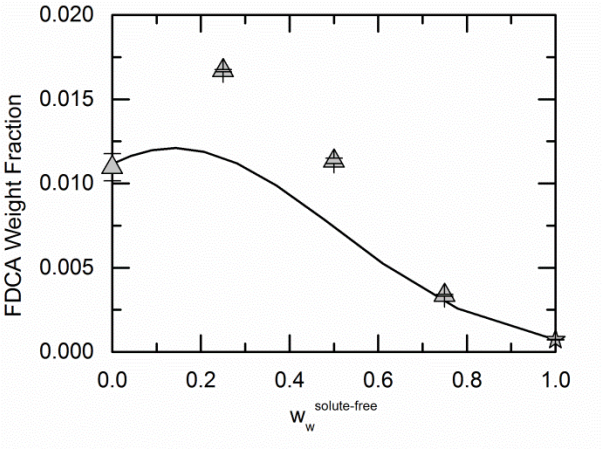


Figure 29: Solid solubility of FDCA in ethanol/water mixtures vs. solute-free water mass fraction in ethanol/water at 35 °C and 1 bar. △: Experimental total solubility data from this work; ☆: experimental solubility of non-dissociated FDCA; lines represent PC-SAFT predictions using the parameters in Table 11 and Table 12.

Although water acts as an anti-solvent for FDCA, a maximum of FDCA solid solubility can be observed upon adding water to ethanol solutions. This is a well-known phenomenon, and has been shown for solid solubility of amino acids or pharmaceuticals in water + cosolvent mixtures. [168] The solid solubility maximum was measured at a solute-free water mass fraction of about 0.25. The occurrence of the solid solubility maximum was validated further with PC-SAFT. PC-SAFT was applied predictively without fitting parameters to FDCA solubility data in the water/ethanol mixtures. Figure 29 illustrates that the PC-SAFT predicted solid solubility is in qualitative agreement with the experimental data, although PC-SAFT slightly underestimates the experimental maximum solubility value. The reason for the quantitative mismatch of the experimental solubility data is probably related to the stated assumption that dissociation does not occur when ethanol is present. As the influence of ethanol on the pK_a values of FDCA is unknown, application of Equation (12) is not feasible.

Further, the influence of ethanol on the solid solubility of FFA in water was investigated. Like FDCA, FFA also has a solid solubility maximum in ethanol/water mixtures at 35 °C, which was found to be at FFA-free water mass fraction of 0.25. This solid solubility behavior was predicted qualitatively with PC-SAFT (results not shown graphically).

4.2.5.6 Mutual Influences on Solid Solubility of Furancarboxylic Acids in Water

Figure 30 and Figure 31 show the mutual influences of FFA and either FA or FDCA on their solubilities in water at 35 °C and 1 bar. Although two similar furancarboxylic acids

are added to saturated FFA+water mixtures, a very different phase behavior can be observed for each system.

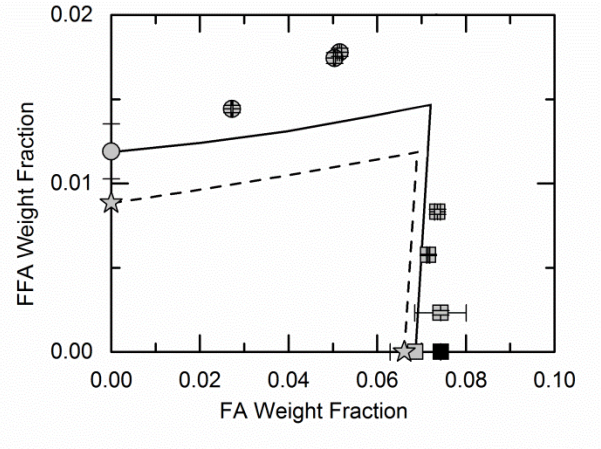


Figure 30: Mutual influence on solid solubility of FFA and FA in water at 35 °C and 1 bar. Symbols are experimental data from this work (○, saturated with FFA; □, saturated with FA; ■, saturated with FA and interpolated from Apelblat and Manzurola (1989) [107]; ☆, solubility of non-dissociated solute in binary cosolute-free system), dashed lines are PC-SAFT predictions of non-dissociated acid $x^{L}_{HA,\pm}$, solid lines are predictions incorporating Equation (12) with pH-values from Table 22 and Table 24. PC-SAFT predictions were made with parameters in Table 11 Table 12.

Figure 30 illustrates that both FFA and FA each increase the solubility of the other in water. This can be predicted in good agreement with PC-SAFT. This prediction includes two effects. The first one is the influence of one component on the solubility of the neutral acid, which is presented as dashed lines. This effect was predicted by Equation (4). It can be observed that this influence significantly contributes to the experimentally observed solubility increase. The second effect is due to the influence of pH on the distribution of protonation states for the acids, i.e. the solubility increases due to higher pH values, which

are caused by adding one acid to a saturated solution of the other acid. These pH values of the ternary solutions water/acid1/acid2 saturated with one of the acids were measured in this work (Table 24). These pH data were used for calculating the solubility curves in Figure 30-Figure 33 using Equation (12). Accounting for the influence of pH shifts each curve parallel to a greater solubility, but pH alone is not sufficient to explain the experimentally-observed solubility increase. It can be concluded from this finding that the mutual influence on solubility is dominated by intermolecular interactions in the liquid phase. Nevertheless, the latter are slightly under-predicted with PC-SAFT.

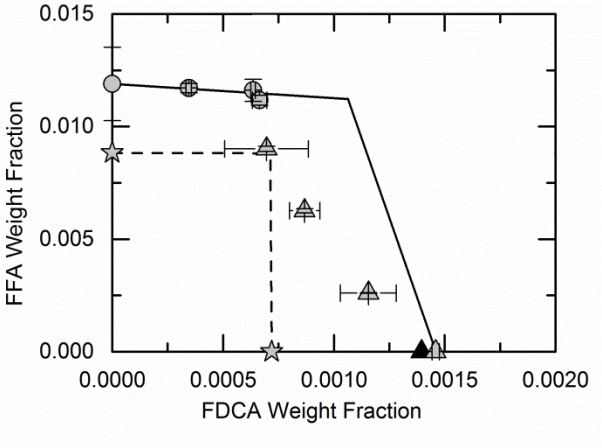


Figure 31: Mutual influence on solid solubility of FFA and FDCA in water at 35 °C and 1 bar. Symbols are experimental data from this work (○, saturated with FFA; △, saturated with FDCA; ▲, saturated with FDCA and interpolated from Lilga, et al. (2010) [144]; ☆, solubility of non-dissociated solute in binary cosolute-free system), dashed lines are PC-SAFT predictions of non-dissociated acid $x^{L}_{HA.\pm}$, solid lines are predictions incorporating Equation (12) with pH-values from Table 22 and Table 24. PC-SAFT predictions were made with parameters in Table 11 and Table 12.

Figure 31 illustrates the mutual influences of FFA and FDCA on their solid solubilities in water. A decrease in the solubility of each solute with an increase in concentration of the other can be observed from the experimental data. Together, PC-SAFT

and pH effects predict this behavior – quantitatively for FFA, while the behavior of FDCA is captured qualitatively but is underpredicted. Despite this quantitative mismatch between data and PC-SAFT prediction, it can be observed from Figure 31 that the concentration of non-dissociated FFA or non-dissociated FDCA does not change with varying cosolute concentration, an effect which was predicted with Equation (4). That is, the molecular interactions between FFA and FDCA play a minor role in these aqueous solutions, whereas pH effects (calculated with pH values from Table 22 and Table 24 using Equation (12)) dominate the phase behavior.

Nevertheless, successfully modeling the solid-liquid phase behavior of the two-solute aqueous system requires considering molecular interactions as well as pH effects. The combination of both allows qualitatively correct predictions using PC-SAFT parameters that were fitted to solubility behavior of exclusively binary single-solute + single-solvent systems.

4.2.5.7 Mutual Influence on Solubilities of a Furancarboxylic Acid and Furfural in Water

Figure 32 and Figure 33 show the solubility of furfural in the presence of the furancarboxylic acids FA and FDCA, respectively, in water at 35 °C and 1 bar. Both figures illustrate that the liquid solubility of furfural is not influenced by the addition of either acid. In contrast, the solid solubilities of FA and FDCA significantly increase upon the addition of furfural. Further, the solid solubility of FDCA is significantly increased by pH effects, whereas the solid solubility of FA is not. One reason for this difference between

these two systems is that FDCA is less soluble than FA by an order of magnitude; thus, the significant dissociation of FDCA does not depend on the presence of furfural, while the concentration of FA is much higher than FDCA, and thus, FA stays mostly in its protonated state independent of the presence of furfural. Based on this finding, molecular interactions between furfural and FA are dominant, as it is shown in Figure 32 that pH contributions to the solubility changes play a minor role.

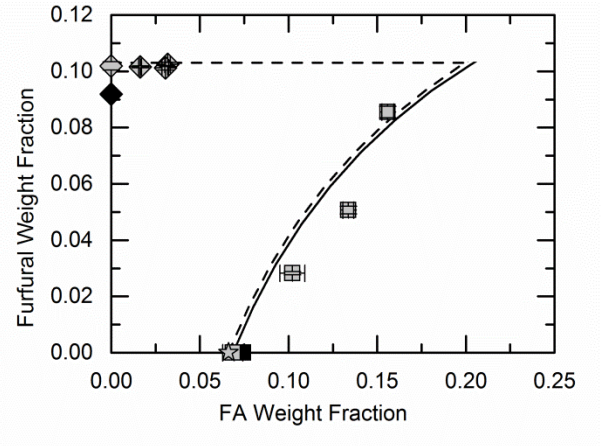


Figure 32: Mutual influence on solubility of furfural and FA in water at 35 °C and 1 bar. Symbols are experimental data from this work (◇, saturated with furfural; ◆, saturated with furfural and interpolated from Mains (1922) [139]; □, saturated with FA; ■, saturated with FA and interpolated from Apelblat and Manzurola (1989) [107]; ☆, solubility of non-dissociated solute in binary cosolute-free system), dashed lines are PC-SAFT modeling results of non-dissociated species, solid lines are predictions incorporating Equation (12) with pH-values from Table 22 and Table 24. PC-SAFT modeling was performed with parameters in Table 11 and Table 12.

However, pH contributes significantly to the solubility increase of FDCA, as illustrated in Figure 33. To model this behavior qualitatively with PC-SAFT one additional binary interaction parameter between furfural and the furancarboxylic acids was applied, which was found to be $k_{ij}(furfural - acid) = -0.28$. A similar solubility behavior was

observed in the literature for aqueous mixtures containing acidic amino acids and neutral amino acids. [151] For such mixtures, very negative k_{ij} values between the two solutes are required for quantitative modeling.

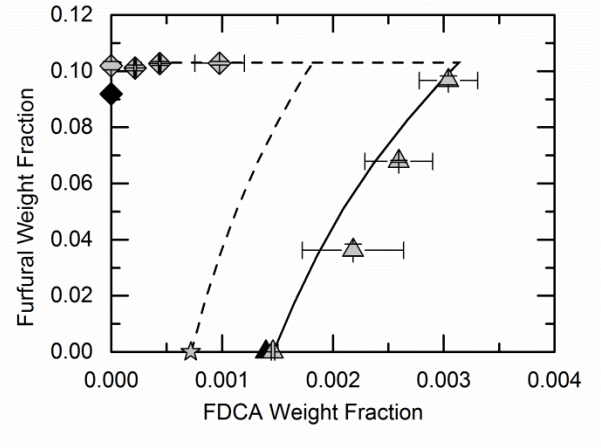


Figure 33: Mutual influence on solubility of furfural and FDCA in water at 35 °C and 1 bar. Symbols are experimental data from this work (◇, saturated with furfural; ◆, saturated with furfural and interpolated from Mains (1922) [139]; △, saturated with FDCA; ▲, saturated with FDCA and interpolated from Lilga, et al. (2010) [144]; ☆, solubility of non-dissociated solute in binary cosolute-free system), dashed lines are PC-SAFT modeling results of non-dissociated species, solid lines are predictions incorporating Equation (12) with pH-values from Table 22 and Table 24. PC-SAFT modeling was performed with parameters in Table 11 and Table 12.

4.2.6 Conclusions and Discussion

The solubilities of FA, FFA, FDCA and furfural in aqueous solution were investigated experimentally and by PC-SAFT modeling. The required PC-SAFT pure-component parameters of the solutes FA, FFA, and FDCA were adjusted to experimental solubility data of each acid in organic solvent. One binary interaction parameter between each acid and water was then adjusted to new solid solubility data of each acid in water. pK_a values allowed predicting the influence of pH on the solubilities of each

furancarboxylic acid in water; these results were validated with experimental solid solubility data measured in this work.

Using the so-obtained PC-SAFT parameters, the solubility behavior of each furancarboxylic acid in multi-component solutions was predicted with PC-SAFT. It was shown that the influence of the solvent ethanol on the solid solubility of FFA or FDCA in water was predicted qualitatively without accounting for pH effects. In contrast, combining pH effects and molecular interactions allowed reasonably predicting the mutual influences of two furancarboxylic acids on their solid solubilities in water. Further, the mutual influences of furfural and furancarboxylic acid on their solubilities in water were accurately modeled applying one binary parameter between furancarboxylic acid and furfural. This further validated all PC-SAFT parameters determined in this work.

In sum, a new body of experimental data and PC-SAFT predictions allowed deconvoluting the influences of pH and intermolecular interactions on the mutual influences of furfural and three furancarboxylic acids on their solubilities. This data and new predictive capability may serve to aid in the development of new reaction systems and separation processes involving the furan series compounds studied in this work.

CHAPTER 5. CALCULATIONS FOR AND APPLICATIONS OF PROTONATION EQUILIBRIA

5.1 Introduction

The thermodynamic modeling performed in the previous chapter provides a strong foundation for handling the intrinsic solubilities of the furan series compounds, and for capturing the associative and dispersion interactions of those species in aqueous and organic solution. However, the work of the previous chapter is limited in handling of pH. Equation (12) captures some aspects of how weak acids dissociate in solution, but it cannot tell a researcher what the pH of a solution will be. The species in an aqueous solution do influence pH, and pH clearly holds influence over the charge state and thus total solubility of weak acids.

The following two subsections, 5.2 and 5.3, explore related methods for calculating the equilibrium pH of aqueous buffer solutions, and showcase these methods with useful applications. The first subsection: A High-Throughput pH-Based Colorimetric Assay: Application Focus on Alpha/Beta Hydrolases, provides a general mathematical strategy for tracking the relationships between equilibrium pH and transient conversion of reactions that produce or consume weak acids. The work relates to hydrolysis and synthesis reactions, however the mathematics of pH are not reaction-specific and the approach is generalizable to other systems, such as the production of FFA from furfural, and the production of FDCA from FFA. The second subsection: Calculation of Ionic Equilibria

and Determination of Acid Dissociation Constants by Isothermal Titration Calorimetry, refines the mathematical tools for the description of pH-related phenomena. This method is extensible to transiently operated continuously-stirred tank reactors, and could even be adapted to predict how heating and cooling duty for buffered reaction systems may evolve over the course of a reaction or chemical process.

5.2 A High-Throughput pH-Based Colorimetric Assay: Application Focus on Alpha/Beta Hydrolases

5.2.1 Statement of Authorship

The text in this section was coauthored by the following: Mariétou F. Paye, [a] Harrison Bellow Rose, [b] John M. Robbins, [b] Diana A. Yunda, [a, c] Seonggeon Cho, [c] and Andreas S. Bommarius. [a, b]

[a] School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States. [b] School of Chemical & Biomolecular Engineering Georgia Institute of Technology, Atlanta, Georgia 30332, United States. [c] School of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States.

All experimental work was performed by Mariétou F. Paye, with assistance from undergraduate research assistants Diana Yunda and Seonggeon Cho. Interpretation of enzymatic data was assisted by research scientist John Robbins. All pH and colorimetry-related modeling and statistical analyses were performed by Harrison B. Rose.

Harrison B. Rose and Mariétou F. Paye contributed equally to this work.

5.2.2 Abstract

Research involving α/β hydrolases, including α-amino acid ester hydrolase and cocaine esterase, has been limited by the lack of an online high throughput screening assay. The development of a high throughput screening assay capable of detecting α/β hydrolase activity toward specific substrates and/or chemical reactions (e.g., hydrolysis in lieu of amidase activity and/or synthesis instead of thioesterase activity) is of interest in a broad set of scientific questions and applications. Here we present a general framework for pH-based colorimetric assays, as well as the mathematical considerations necessary to estimate *de novo* the experimental response required to assign a 'hit' or a 'miss,' in the absence of experimental standard curves. This combination is valuable for screening the hydrolysis and synthesis activity of α/β hydrolases on a variety of substrates, and produces data comparable the current standard technique involving High Performance Liquid Chromatography (HPLC). In contrast to HPLC, this assay enables improved efficiency obtained from a single experiment.

5.2.3 Introduction

5.2.3.1 AEH and CocE Enzymes

Hydrolases are ubiquitous enzymes involved in a myriad of chemical processes and are essential to the survival of all living organisms. [169] The most widely studied enzymes in this class are in the α/β hydrolase family, which have applications in analytical, industrial, and pharmaceutical fields. Some uses include synthesis of β-lactam antibiotics (β-LA) with α amino acid ester hydrolase (AEH, Figure 34 A), [170] treatment of cocaine

addiction with cocaine esterase (CocE, Figure 34 B), [171] and hydrolysis of insecticides by organophosphorus hydrolase (reaction scheme not shown). [172]

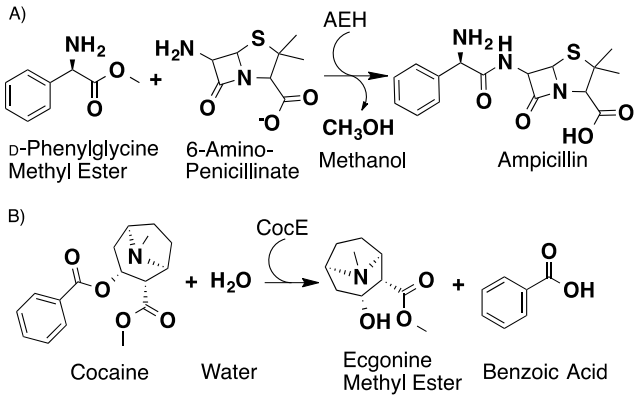


Figure 34: Popular applications of the α-amino acid ester hydrolase (AEH) and cocaine esterase (CocE) enzymes used in this study. A) AEH catalyzes the synthesis of ampicillin from D-phenylgylycine methyl ester and 6-aminopenicillinate B) CocE catalyzes the hydrolysis of cocaine into benzoic acid and ecgonine methyl ester.

There have been intensive efforts to identify amino acid substitutions and reaction conditions that improve thermostability and catalytic rates of AEH and CocE. [171, 173-174] However, those efforts have been limited by dependence on HPLC-based assays, which are low throughput and resource-intensive. Thus, an alternative, high-throughput assay is desirable to accelerate the process of generating improved enzyme variants. For this purpose, a model reaction in which AEH or CocE catalyzes the sequential peptide synthesis of multiple D-phenyl glycine methyl ester (D-PGME) molecules, thus resulting in a mixture of various peptide methyl esters, has been proposed (Figure 35). [175] A crucial side-reaction also catalyzed by these enzymes is the primary hydrolysis of D-PGME, which releases D-phenyl glycine (D-PG) and methanol (Figure 35). Because hydrolysis of the

reacting methyl ester limits the amount of active substrate available for the synthesis reaction, biocatalysts capable of promoting the synthesis reaction while suppressing, if not blocking completely, the hydrolysis reaction is coveted by industry. [176] Given the need for a less resource-intensive analytical method, the current study serves to utilize this model reaction to identify a high throughput assay capable of differentiating between the ester hydrolysis and amide synthesis reactions in general.

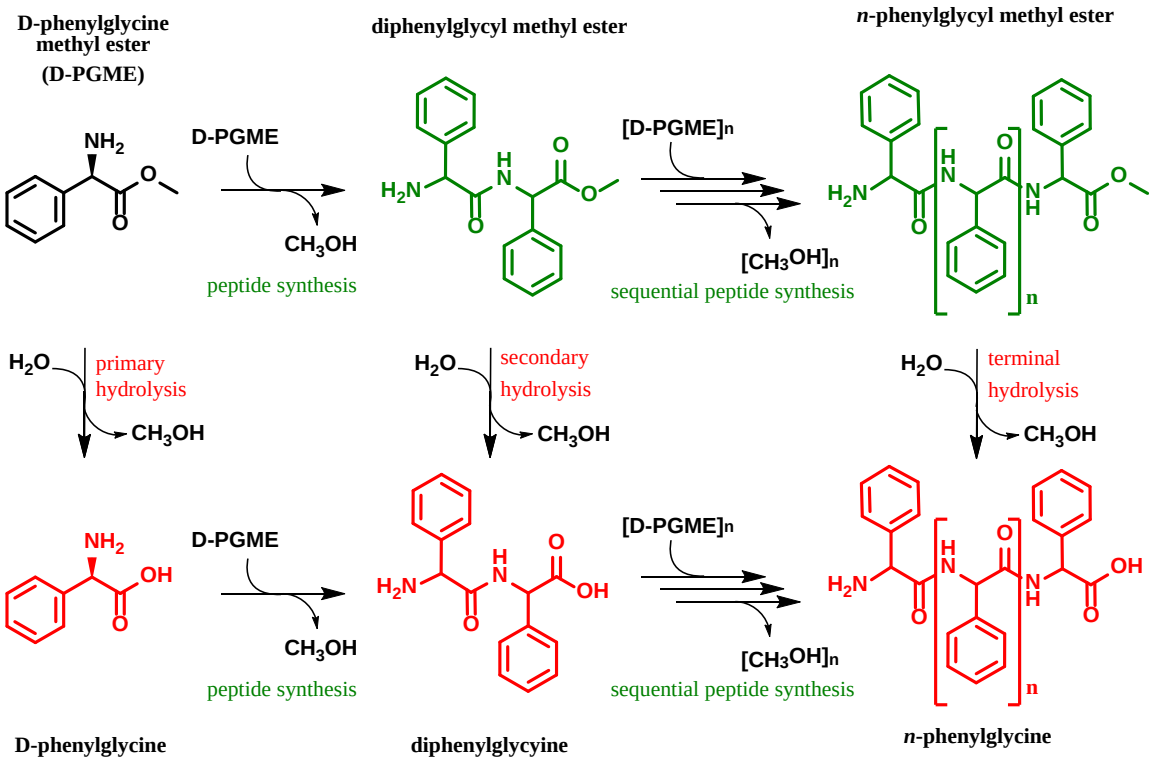


Figure 35: Proposed scheme for the α-amino ester hydrolase (AEH) or cocaine esterase (CocE) catalyzed model reaction of D-PGME to various products. Sequential peptide synthesis reactions (left to right) are proposed to occur in competition with ester hydrolysis reactions (top to bottom).

5.2.3.2 Literature Survey of Existing Assays

Since the substrates and products (i.e. primary amines and carboxylic acids, respectively) of these enzymes have labile protons, it was proposed that weakly buffered systems may experience a change in pH value over the course of ester hydrolysis or amide synthesis reactions such as those shown in Figure 34 and Figure 35. The use of pH-based colorimetry to detect enzyme activity has been reported with certain hydrolases (enantioselective hydrolases), [177] transketolases (e.g., TK, EC 2.2.1.1), [178] transferases (e.g., galactosyltransferase), [179] acylases (e.g. cephalosporin C acylase), [180] and kinases (e.g., glucokinase). [181] These published assays are dependent on low buffering capacity and salt content, selection of pH-indicator based on initial pH (the pK_a of the indicator should equal the starting pH), and high concentrations of pH-indicators. Additionally, each set of conditions requires unique standard curves, which are time-consuming to collect. Collectively, these restrictions limit the extent to which various substrates and reaction conditions can be explored, and the rapidity with which such investigations can be conducted.

Two colorimetric assays for AEH activity have been described previously, one endpoint assay, and one for real-time monitoring (see Figure 59 in APPENDIX C, section C.1.2, Other Assays for Alpha-Beta Hydrolases). The endpoint assay is designed to measure the absorbance of the chromophore diketopiperazine (DKP) derivative, which is formed by ring-opening of β-LA coupled with nucleophilic attack by reduced ascorbic acid [182] However, the reaction does not effectively distinguish between substrates and products. [183] On the other hand, the real-time assay tracks the production of the

chromophore 5-amino-2-nitrobenzoic acid, which is produced by hydrolysis of an AEH substrate analog, 5-(2'-amino-5-phenylacetamido)-2-nitrobenzoic acid (APANiB).[183] APANiB is not commercially available and is difficult to synthesize due to its precursors' tendency to self-polymerize.

Similarly, a real-time CocE assay using substrate analogs has been reported.[184] In this assay scheme (illustrated in Figure 59 in APPENDIX C, section C.1.2, Other Assays for Alpha-Beta Hydrolases), enzymatic cleavage of the thioester bond in benzoyl-3-mercapto ecgonine methyl ester produces benzoate and 3-mercapto ecgonine methyl ester, which is reacted with 5-5'-dithiobis(2-nitrobenzoic acid) (DTNB, Ellman's reagent) to produce 3-5'-dithio(ecgonine methyl ester)-(2-nitrobenzoic acid), and the chromophore, 5-mercapto-2-nitrobenzoic acid.[184] However, this assay is limited to detecting thioesterase activity and requires the use of a thioester analog of each target substrate.

5.2.3.3 Generalizing pH-Based Colorimetric Assays

The development of an improved assay method was motivated by the challenges surrounding current AEH and CocE assays, as well as pH-based assays in general. In reaction theory, conversion is a single value that encapsulates the multitude of variables related to the progress of a reaction or, in an assay context, the success of a catalyst. We present a general mathematical approach for creating a direct mapping between the absorbance of a colorimetric indicator and the conversion of a pH-altering reaction. Our model quantitatively predicts how an assay will respond to a reaction, eliminating the need for experimental standard curves in 'hit' or 'miss' activity assessment and reducing the

time and material resources needed to conduct high throughput screens, and even enabling real-time monitoring and hypothesis testing. Assay response curves are typically unintuitive – changes in solution pH tend to be accompanied by changes to ionic strength which in turn alter the effective pK_a values of participating species and may broaden, narrow, or shift the buffering capacity of the reaction mixture. By allowing operators to visualize the response and sensitivity of an assay to each design parameter, these tools can also serve to support and inform decisions made in assay development and troubleshooting. We present some examples of general experimental guidelines and troubleshooting recommendations that have been borne out by the modeling in this manner.

For aqueous hydrolysis reactions of known stoichiometry, it is possible to estimate the degree of conversion that is required to produce a particular pH value, so long as the starting concentrations and acidity constants of all species are known. Generally, pH can also be related to the color of a halochromic indicator. Phenol red was selected as the indicator for this work, as it exhibits a visible color change over the pH range (4.5-8.5) of interest for monitoring hydrolysis and synthesis reactions. By combining these relationships, it is possible to predict the spectrophotometric response of a hydrolysis reaction. Plots of absorbance vs. hydrolysis conversion produced with this method can be used to identify assay conditions that provide reasonable sensitivity and to rule out specific reactions for which pH-based assays are inappropriate. Once experimental absorption data (either endpoint or real-time) are collected, the predicted responses may be used in lieu of experimental standard curves for data interpretation and rapid qualification of “hits” and “misses.”

In the sections that follow, we validate and demonstrate this approach to pH-based colorimetric assays in several steps. First, we compare enzymatic reaction data obtained by high performance liquid chromatography (HPLC) using *traditional* standard curves (i.e., for quantification of individual species) with data obtained by colorimetry using *simulated* standard curves (i.e. solutions that have been deliberately prepared to simulate the state of a reaction system at a given level of conversion). Next, we compare *predicted* standard curves (i.e., calculated by the model without fitting) to those *simulated* standard curves. Finally, we use the *predicted* standard curves from the model to interpret experimental enzymatic reaction data and perform hypothesis testing to demonstrate "hit" or "miss" screening. The proposed workflow for an assay experiment with hypothesis testing is illustrated in Figure 36 (and in more detail in Figure 58 in APPENDIX C, section C.1.1, Workflow and Dependency of Steps).

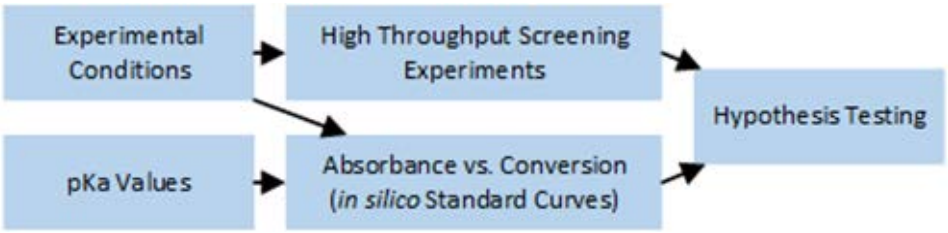


Figure 36: The workflow required to use the generalizable pH-based colorimetric assay. The experimental conditions used in wet lab experimentation are tabulated and combined with existing knowledge of substrate, product, and buffer pKa values, which are used to predict the relationship between spectrophotometric measurements and reaction conversion. Experimental data and predicted relationships are then combined to perform hypothesis testing and qualify each experiment as either a hit or a miss.

5.2.4 Methods

5.2.4.1 Materials

Sodium dodecyl sulfate, nicotinamide, D-phenylglycine methyl ester, phenol red, and methyl picolinate were purchased from Sigma Aldrich while acetamide, nicotinic acid, imidazole, and picolinic acid were purchased from Alpha Aesar. Other commercial sources include Fisher Scientific for ammonium acetate, sodium hydroxide, and HPLC-grade acetonitrile; Fluka Chemika for 6-aminopenicillanic acid and D-phenylglycine; VWR analytical for dibasic anhydrous sodium phosphate and sodium chloride. Other purchased materials were sodium ampicillin (Gold Biotechnology), Overnight Express (Novagen), monobasic sodium phosphate monohydrate (J.T. Baker), molecular biology grade LB broth lennox (US Biological Life Sciences) and hydrochloric acid (Acros Organics). The expression vector for α-amino acid ester hydrolase (AEH) was produced by a past member of this research group [175, 185] while that for cocaine esterase was graciously donated by Dr. Neil Bruce's laboratory at the University of York, UK.

5.2.4.2 Protein Expression and Purification

Both AEH and CocE were expressed in Rosetta™(DE3)pLysS competent cells (Novagen) using pET28 and p234 under selective pressure from kanamycin and ampicillin-carbenicillin, respectively. Because wildtype AEH is not thermostable, only its quadruple variants H (N186D/A275P/V622I/E143H: AEH QV-H) and G (N186D/A275P/V622I/E143G: AEH QV-G) were used in this paper. [185] Chloramphenicol was used to maintain the plasmid supplying tRNAs for exogenous codons. 5 mL seed

cultures (LB broth, 37 °C, 200 rpm) were grown overnight, passaged to 200 mL of Overnight Express, and incubated for an additional 8 hours (37 °C, 250 rpm). Cultures were incubated at room temperature for 16 hours after which they were centrifuged. The resulting pellets were suspended in lysis buffer (18 mL per 1 g of pellet; 10 mM Imidazole, 50 mM $NaPO_4$, 300 mM NaCl, pH 8.0) and lysed by sonication on ice. Enzymes were purified by nickel affinity chromatography with Novagen Ni-NTA HisBind Resin. Bound proteins were washed (1 mL of 50 mM Imidazole, 50 mM $NaPO_4$, 300 mM NaCl, pH 8.0; repeated 20x), followed by 1 mL washes with buffers containing imidazole increasing in concentration from 60 mM to 240 mM in 20 mM increments (remaining buffer composition: 50 mM $NaPO_4$, 300 mM NaCl, pH 8.0). Washed proteins were eluted with 8 mL of elution buffer (250 mM Imidazole, 50 mM $NaPO_4$, 300 mM NaCl, pH 8.0). For each enzyme, the combined fractions were dialyzed twice against 2 L of dialysis buffer (100 mM $NaPO_4$, 100 mM NaCl, pH 7.0) to remove the imidazole, which inhibits enzyme activity. Purified enzyme was quantified using Bradford reagent at the absorbance maxima of 595 nm.

5.2.4.3 Enzymatic Hydrolysis and Synthesis Reactions

Five types of hydrolysis and synthesis reactions were carried in this study: the reactions of (1) 6-aminopenicillinate and D-phenylglycine methyl ester to ampicillin and methanol; (2) methyl picolinate to picolinate and methanol; (3) ampicillin to 6-aminopenicillinate and D-phenylglycine; (4) D-phenylglycine methyl ester to D-phenylglycine and methanol, (5) and D-phenylglycine methyl ester to di- and tri-peptides and methanol.

All reactions were carried out in 96-well clear flat-bottom plates at room temperature in 200 μL of 100 mM sodium phosphate buffer with 20 μM phenol red. Substrates were added to the required concentration, which varied between experiments. Reaction start-time was taken upon addition of purified enzyme, either CocE or AEH. Enzymes were suspended in dialysis buffer and added to the required final concentration. Non-enzymatic controls were prepared using dialysis buffer without enzyme. Additional controls were performed to verify that phenol red is neither a substrate nor an inhibitor of CocE or AEH, including trials with enzyme and substrate but without phenol red, as well as with enzyme and phenol red in the absence of substrate. These controls are described in APPENDIX C, section C.1.4, Validation of Phenol Red.

Reactions were monitored with direct measurements of product formation by high performance liquid chromatography (HPLC) using a Shimadzu HPLC 20A (DGU-20A_5, LC-20AT, SIL-20AC HT, CBM-20A, SPD-M20A, CTO-20A) and a reverse-phase Phenomenex C-18 column (Luna 5 μ C18 (2) 100 Å 250×4.60 mm 5 micron). The mobile phase consisted of 70% 5 mM phosphate buffer at pH 3.0 with 300 mg/L SDS and 30% acetonitrile supplied at a rate of 1 mL/minute. Reaction time points were taken by 10-fold dilution of reaction aliquots into the mobile phase. Diluted samples were then supplied to the column in a 2.0 μL injection. Column eluent was monitored at 215 nm to monitor all substrates and products.

The pH of each reaction mixture is an aggregate result of the relative buffering capacity contributions of the background buffer, the reaction substrate(s) and the reaction product(s). For that reason, the pH value of each sample was indirectly monitored by

tracking changes in the absorbance of phenol red on a Biotek Synergy H4 Hybrid (plate) reader. Absorbance measurements were taken over time at 557 nm and 479 nm, the alkaline absorbance maximum and isosbestic point of phenol red, respectively. All downstream data processing was performed using the (557 nm/479 nm) absorbance ratio. Full wavelength scans from 300 nm to 700 nm were also performed on each well after 4 hours to check for baseline deviation.

5.2.4.4 Simulated Conversion Standard Solutions

Samples were prepared to standardize the spectrophotometer analyses for each hydrolysis and synthesis reaction. Stock solutions of reactants were prepared in 100 mM sodium phosphate. For reasons discussed later in the modeling section, precise accounting of total ions is critical to the success of this method. The sodium concentration was adjusted to give the reactant stocks a pH value of 7.0. Individual product stocks were prepared in a 2x buffer of the same relative salt composition as the reactant stocks, such that mixing the reactant and corresponding product stocks in predefined ratios simulates a desired conversion without changing the buffer composition. For example, mixing a 1:3:3 ratio of the stocks for a reactant and both products generates a solution with the same composition and pH as a reaction at 75% conversion. Each sample was evaluated by spectrophotometry and HPLC as described above.

5.2.4.5 Determination of Enzymatic Kinetic Parameters

Using the simulated conversion standards to interpret the spectrophotometric and HPLC data, initial rates were determined in triplicate for each reaction in both the presence

and absence of enzyme. The contribution of chemical (non-enzymatic) hydrolysis was subtracted, and the kinetic parameters were determined from the best fit of data to the Michaelis-Menten equation using Kaleidagraph [a].

5.2.4.6 Characterization of Phenol Red

In addition to spectrophotometric evaluation of the simulated conversion standards, a 20 μM solution of phenol red in 100 mM phosphate buffer was titrated with sodium hydroxide (10 N) and hydrochloric acid (6 N) to reveal the dependence of its absorption spectrum on pH in the absence of other compounds.

5.2.4.7 Modeling and Prediction of Assay Response Curves

A general mathematical model was developed to predict how the colorimetric assay would respond to different hydrolysis reactions. The predictions are used for experimental planning, and to interpret experimental results for hypothesis testing, as described below.

Hydrolysis reactions examined here are of the general form $A + water \rightleftharpoons B + C$ where A, B, and C may be weakly acidic or basic. Model formulation begins with the general consideration of a weak acid reversibly dissociating in aqueous solution: $AH_j^{z+1} \rightleftharpoons H^+ + AH_{j-1}^{z}$ where z is the charge of the conjugate base and j is the number of labile protons in each protonation state.

[a] Synergy Software, KaleidaGraph V4.1.1, Reading, PA (1986-2000)

The degree to which a weak acid dissociates is typically reported as a pK_a value which may be presented as a pK_a^c ('concentration basis'), pK_a^{th} ('thermodynamic value' or 'activity-basis'), or pK_a^* ('mixed-mode' or 'practical'), depending on the data source and application. [186] [b] pK_avalues determined using a potentiometric pH probe are typically in the form of pK_a^*, unless they have been extrapolated to infinite dilution (pK_a^{th}), or the probe has been standardized using particularly rigorous methods (pK_a^c). [187] Acid-base equilibria inherently involve charged species and are sensitive to (and result in) changes in ionic strength. As a result, deviations between pK_a^c, pK_a^{th}, and pK_a^*, as well as between activity- and concentration-based measures of pH tend to increase with increasing ionic strength.

The concentration-based pK_a value, pK_a^c, may be written as the sum of the thermodynamic value and a 'lumped activity coefficient' correction factor that is dependent on ionic strength. To keep this model as general as possible, an extended Debye-Hückel relationship is used to estimate the correction factor. This correction can be expressed as a special case of the correlation derived by Sun et al: [156]

$$pK_a^c = pK_a^{\mathrm{th}} + (1.02z)\left(\frac{\sqrt{I}}{1 + A\sqrt{I}} + BI\right) - CI \qquad (18)$$

[b] A complete discussion of these bases, with guidelines for standardizing literature values to a thermodynamic basis is provided in the appendix section C.2.2, Standardization of pKₐ Values from the Literature. Additionally, a tool for standardizing and curating pK_a values is provided in the first 'sheet' of the spreadsheet accompanying this work.

where z is the charge of the conjugate base, I is the ionic strength, and $A, B,$ and C are empirically derived parameters with mean values of 1.50, -0.09, and -0.09, respectively. The correction is valid up to ionic strengths of $I = 0.5$ M, and enables the use of concentrations rather than activities elsewhere in this model. A detailed explanation of this correction is provided in APPENDIX C, section C.2.2.

Similarly, pH can be adjusted for ionic strength, to convert between the concentration-based value, pH_c, and the activity-based value, pH_a, with the latter corresponding to readings from a typical potentiometric pH electrode.

$$pH_a = pH_c + 0.51\left(\frac{\sqrt{I}}{1 + 1.50\sqrt{I}} - 0.11I\right) \tag{19}$$

where

$$pH_c = -\log_{10}([H^+]) \tag{20}$$

Finally, these general relationships may be used to determine the distribution of protonation states:

$$pK^{\mathrm{c}}_{a,i,j} = -\log_{10}\left(\frac{\left[AH_{j-1}\right] \times 10^{-pH_c}}{\left[AH_j\right]}\right) \tag{21}$$

for all ionizable species, i, and number of labile protons, j. Note that pH_c is used here to avoid over-correcting for ionic strength.

The protonation state of water can also be represented, using the definition of the ionic product of water:

$$[OH^-] \times 10^{-pK_W^c} = 10^{-pH_c} \tag{22}$$

Here also, pH_c is used. The value of pK_W may be corrected as in Equation 1, however, a value of $z = 1$ must be used.

It is necessary to link these expressions together using a charge balance:

$$0 = \sum_i \sum_j \left(c_{i,j} \times z_{i,j}\right) \tag{23}$$

where $c_{i,j}$ and $z_{i,j}$ are the molar concentration and charge number of each ion in each protonation state, respectively. If the solutions are titrated to a specific pH and/or prepared using salts, then the concentrations of the co-salt ions and titrants must be included in the charge balance equation. Precision is required in the accounting of all salt species. An error in accounting for a monovalent anion on paper, for example, is comparable to the erroneous addition of hydrochloric acid in the laboratory.

Additionally, since specific protonation states cannot be measured or pipetted, total concentrations are used to specify and track the analytical composition of the reaction system:

$$c_{i,tot} = \sum_{j} \left(c_{i,j} \right) \tag{24}$$

This treatment is not necessary for the concentrations of sodium and chloride, as both can be calculated based on solution and stock preparation, and are assumed to be fully dissociated.

To predict the spectrophotometric response of the system due to the colorimetric indicator, the ratio of the molar extinction coefficient at one select wavelength relative to that of the isosbestic point is calculated for each protonation state of the indicator:

$$Abs_j = \frac{Abs_{j,\lambda_{sensitive}}}{Abs_{\lambda_{isosbestic}}} \tag{25}$$

When phenol red is used as the assay indicator, we use 557nm for $\lambda_{sensitive}$ and 479 nm for $\lambda_{isosbestic}$.

Next, the sum of the absorbance ratios of all protonation states of the indicator is normalized by the total indicator concentration:

$$Abs = \frac{\sum_j \left(Abs_j \times c_{indicator,j} \right)}{C_{indicator,tot}} \tag{26}$$

It is also necessary to define the stoichiometric relationships between the reacting species in a manner that considers the conversion of the enzymatic reaction, and accounts for the possibility of non-zero starting concentrations of the products:

$$[A_{tot}] = [A_{init}] \times (1 - X) \tag{27}$$

$$[B_{tot}] = [B_{init}] + [A_{init}] \times X \tag{28}$$

$$[C_{tot}] = [C_{init}] + [A_{init}] \times X \tag{29}$$

The inclusion of the reaction in this manner relies on the assumption that the acid-base equilibrium reactions are significantly faster than the enzymatic reaction for any level of reaction conversion.

The ionic strength of the solution is dependent on the distribution of charged species present, and is calculated as follows:

$$I = \frac{1}{2}\sum_{i}\sum_{j}\left(c_{i,j} \times z_{i,j}^{2}\right) \tag{30}$$

Relating the total absorbance of the reacting mixture to its conversion requires solving the system of Equations (18)-(30). These equations are written in their expanded forms with additional explanation in APPENDIX C, section C.2.4, Explicit System of Model Equations.

In general, models of this form can be difficult to solve numerically, as the species concentrations spanning several orders of magnitude makes them poorly scaled. It is also difficult to solve such systems analytically for pH as a function of the species distribution. As such, approximations and simplifying techniques are frequently deemed necessary to make even seemingly straightforward buffer calculations (such as those without ionic strength corrections or concurrent hydrolysis reactions) more mathematically tractable. [188]

However, by treating pH_c as the independent variable, it is possible to solve the system of Equations (20)-(30) algebraically. Solution sets are defined by fixed parameters, including the initial assay composition ($[A]_{init}$, $[B]_{init}$, $[C]_{init}$, $[Phos]_{tot}$, $[Indicator]_{tot}$, $[Na^+]$, and $[Cl^-]$), thermodynamic ionization constants $pK_{a,i,j}^{th}$, and indicator absorptivities Abs_j. Wolfram Mathematica [c] was used to solve Equations (20)-(30) symbolically, producing algebraic expressions for ionic strength, absorption, and conversion as functions of pH_c. [d] These expressions are embedded within a MATLAB [e] framework, which allows them to be easily evaluated for a variety of reactions and reaction conditions which are specified within an accompanying spreadsheet.

In each execution of the MATLAB code, the complete expression for ionic strength is evaluated over a range of pH_c values, using the thermodynamic values for each acid

[c] Wolfram Research, Inc., Mathematica, Version 10.4, Champaign, IL (2016)

[d] The Mathematica inputs and outputs are provided in the appendix, in sections C.2.5, Mathematica Input: Solving the Model Equations and C.2.6, Mathematica Output: Expressions for Absorbance, Conversion, and Ionic Strength, respectively, as well as in the accompanying Mathematica Notebook file.

[e] MATLAB Release 2016a, The MathWorks, Inc., Natick, MA.

dissociation constant. The calculated ionic strengths are then used in Equation (18) to estimate the concentration-based pK_a^c value of each species over the pH_c range, and the ionic strength is calculated again, using the new pK_a^c values. Three such iterations are typically sufficient for reasonable convergence. The resulting pK_a^c values are then used in the evaluation of the complete expressions for absorbance and conversion at each pH_c value. Finally, pH_a is calculated as described in Equation (19), so that the pH axis of the displayed data plots will correspond to traditional potentiometric measurements. A guide to using the accompanying MATLAB script is provided in APPENDIX C, section C.2.1, Modeling Quick-Start Guide, and a more detailed explanation can be found in APPENDIX C, section C.2.7, MATLAB Script Explanation.

As it is unlikely that perfect information will be available when parameterizing the model, the model framework has been structured to draw parameters from distributions mimicking the uncertainty of the experimental information. When using this Monte Carlo approach, reasonable estimates of uncertainty in model parameters (standard deviations of ± 0.1 on pK_a^{th} values and $\pm 2.5\,\%$ on nonzero concentrations) are sufficient to encompass most of the experimental simulated conversi on data. Implementation of Monte Carlo uncertainty propagation and sensitivity analysis in the model are discussed in more detail in APPENDIX C, section C.2.9, Model Resolution, Sensitivity, and Error Analysis.

This model was used to predict assay response curves that might be observed under various conditions for each of the hydrolysis reactions considered, as well as to estimate the uncertainties associated with each prediction. Each prediction was initialized using a

spreadsheet containing the model parameters, including the initial experimental conditions of the assay reaction, as well as the charges and pK_a^{th} values of each of the ionizable species. An example spreadsheet for initializing the model is also included, and an explanation is provided in APPENDIX C, section C.2.3, Excel Spreadsheet. The charges and ionization constants used in this study are listed in Table 27 in APPENDIX C, section C.2.2.2, Charges, Protonation States, and Curated pK_a Values for Model Species.

5.2.4.8 Hypothesis Testing

The predicted relationship between reaction conversion and absorbance ratio can be used to interpret experimentally measured absorbance ratios as reaction conversions, with appropriate levels of uncertainty. This calculation can be performed for both enzymatic and non-enzymatic experiments, and their interpreted conversion levels can then be compared to determine if the presence of an enzyme has any effect on reaction conversion. This comparison requires the formation of two hypotheses: H_0 (the null hypothesis), which states that the enzyme has no impact on reaction conversion (an assay 'miss'), and H_1 (the alternative hypothesis), which states that the enzyme does impact the reaction (an assay 'hit'). The 'Welch's t-test,' using the Welch-Satterthwaite equation to estimate the underlying degrees of freedom, is applied to determine the probability that the enzymatic and non-enzymatic conversion values are equal. [189] Finally, using any desired p-value threshold α, the null hypothesis H_0 is either accepted or rejected. This analysis can be performed with a single set of experimental measurements or with time-series kinetic data, and is performed automatically with the supplied spreadsheet.

5.2.5 Results

5.2.5.1 Characterization and Validation of Phenol Red

The color change of phenol red over a range of pH values from 6.2 to 8.1 is presented in Figure 37. Full absorption spectra of phenol red under varied conditions and over a broader range of pH values are provided in Figure 61 in APPENDIX C, section C.1.4, Validation of Phenol Red. After baseline removal, spectrophotometric analyses of reactions were performed by normalizing (see Figure 62 in APPENDIX C, section C.1.4, Validation of Phenol Red) the absorbance at 557 nm by that at the isosbestic point, 479 nm, and nonlinear regression was performed in MATLAB (Figure 63 in APPENDIX C, section C.1.4, Validation of Phenol Red) to parameterize the model. No change in the absorbance of phenol red was detected in the presence or absence of CocE (Figure 64 in APPENDIX C, section C.1.4, Validation of Phenol Red), and HPLC analysis of reaction controls revealed no change in the activities of either CocE or AEH in the presence or absence of phenol red (Figure 65 in APPENDIX C, section C.1.4, Validation of Phenol Red). These results suggest that phenol red is neither a substrate nor an inhibitor of the enzymes used in this study.

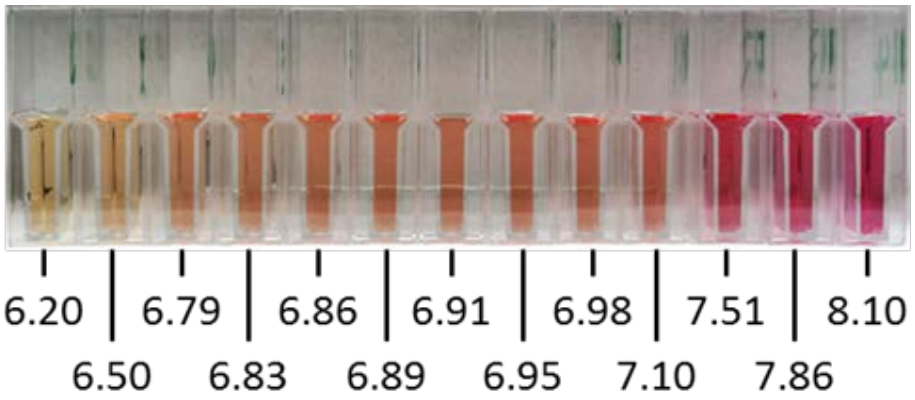


Figure 37: The color change of phenol red. Each cuvette contains 20 µM phenol red in 100 mM sodium phosphate over a range of pH_a values from 6.20 to 8.10. This broad range of pH sensitivity makes phenol red a suitable indicator for assaying a wide variety of hydrolysis and synthesis reactions.

5.2.5.2 Evaluation of Hydrolysis Activity by Colorimetry and HPLC

To validate the colorimetric assay, both HPLC and spectrophotometry were used to evaluate enzymatic hydrolysis activity and the results were compared. Two reactions were tested: the hydrolysis of methyl picolinate by CocE (Figure 38), and the hydrolysis of ampicillin by AEH variant QV-G.

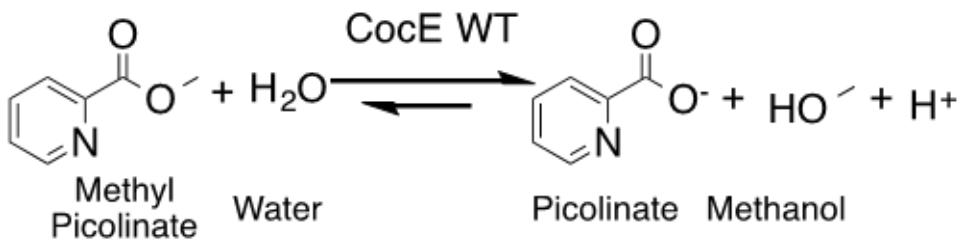


Figure 38: Cocaine Esterase (CocE) catalyzes the hydrolysis of methyl picolinate, producing picolinate and methanol. The carboxylic acid group on picolinic acid will release a proton under basic or neutral conditions, allowing the progress of this reaction to be tracked by indirect (spectrophotometric) monitoring of the solution pH using phenol red.

Experimental data and kinetic parameters for the hydrolysis of methyl picolinate by CocE are presented in Table 13, while experimental data and kinetic parameters for the hydrolysis of ampicillin by AEH variant QV-G are presented in Figure 67 in APPENDIX C, section C.1.5, Enzyme Kinetics. Plots of CocE (0.148 μM) hydrolysis activity monitored by both colorimetric and HPLC methods as a function of methyl picolinate concentration are shown in Figure 39, and fitted kinetic parameters are reported in Table 13 (±SE). Parameter values determined by colorimetry ($K_M = 17 \pm 6$ mM, $V_{max} = 230 \pm 20$ pmole/s, $k_{cat} = 7.9 \pm 0.6\, M^{-1}s^{-1}$, [CocE] = 0.148 μM) were consistent with those determined by HPLC ($K_M = 14 \pm 7$ mM, $V_{max} = 200 \pm 20$ pmole/s, $k_{cat} = 6.6 \pm 0.7\, M^{-1}s^{-1}$, [CocE] = 0.148 μM). Additionally, values of V_{max} were shown to increase proportionally with enzyme concentration, further confirming the validity and sensitivity of the colorimetric assay. Additional kinetic data are provided in Figure 66 in APPENDIX C, section C.1.5, Enzyme Kinetics.

Table 13: Apparent net kinetic parameters for hydrolysis of methyl picolinate by CocE.

Instrumentation	No.	[CocE] (μM)	K_M (mM)	V_{max} (nmol s^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
HPLC	1	0.148	14 ± 7	200 ± 20	6.6 ± 0.7	470 ± 50
Spectrophotometer	2	0.123	17 ± 6	200 ± 20	8.3 ± 0.7	490 ± 40
Spectrophotometer	3	0.148	17 ± 6	230 ± 20	7.9 ± 0.6	460 ± 40
Spectrophotometer	4	0.177	17 ± 6	270 ± 20	7.5 ± 0.6	440 ± 30

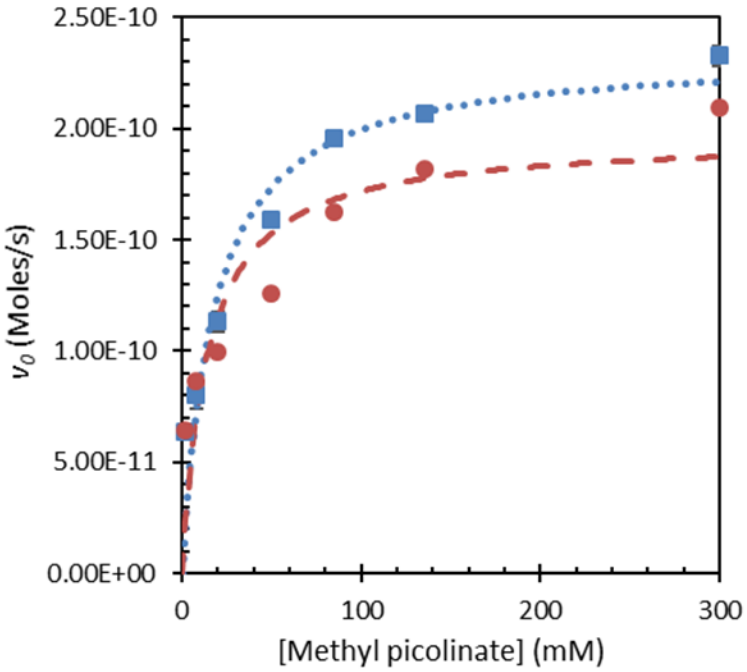


Figure 39: Dependence of methyl picolinate hydrolysis rate on substrate concentration. Data points represent steady-state kinetics of the hydrolysis of methyl picolinate to picolinate and methanol by wild-type cocaine esterase (CocE) (0.148 μM), as determined by spectrophotometry (■) and by HPLC (●). Curves represent fits of the Michaelis-Menten equation to the data collected using the spectrophotometric method (·····) and HPLC method (– –).

The ability of the colorimetric assay to capture kinetic parameters for both the hydrolysis of methyl picolinate by AEH and the hydrolysis of ampicillin by AEH was consistent with the liberation of a carboxylic acid causing solution pH to decrease as either reaction progresses. Both analytical methods evaluated here, HPLC and colorimetry, revealed kinetic data consistent with each other and supported the use of pH-based colorimetry as a viable alternative to HPLC.

5.2.5.3 Differentiation Between Synthesis and Hydrolysis Activity

To demonstrate that the colorimetric assay was able differentiate between synthesis and hydrolysis activity, a model reaction with D-PGME serving as the sole substrate was implemented (Figure 35). Because the colorimetric assay was shown to be sensitive to pH, hydrolysis reactions that decreased the pH value of initially neutral solutions could be distinguished from amide bond forming synthesis reactions that increased the pH value of initially neutral solutions. The hydrolysis and synthesis reactions of 25 mM D-PGME by either CocE or a thermally stabilized variant of wild type AEH (AEH QV-H) were evaluated, and the solution pH was indirectly monitored by tracking the absorbance of phenol red at 557 nm in the reaction mixtures (Figure 40).

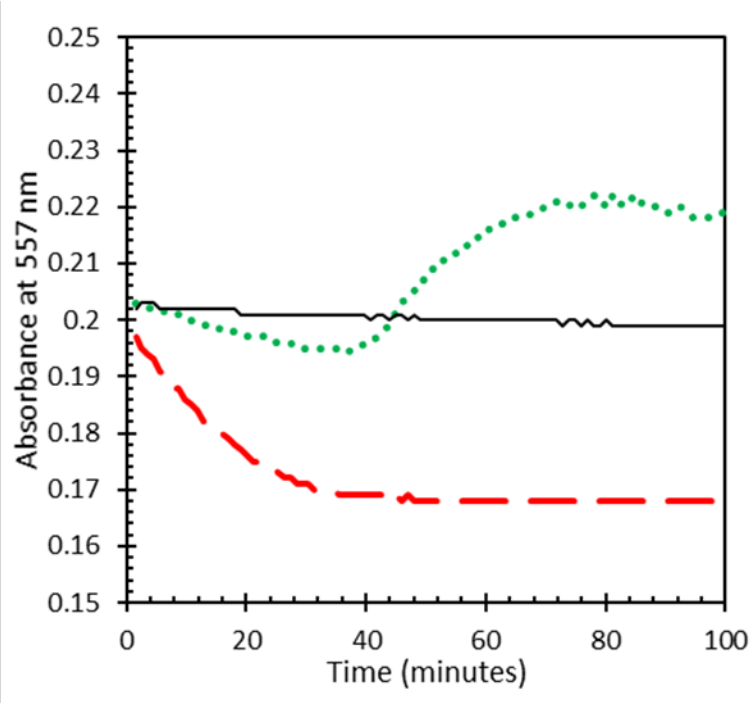


Figure 40: Measured absorbance as D-PGME is catalyzed by AEH and CocE. Each trace shows the absorbance at 557 nm over time for reaction mixtures containing 25 mM D-PGME and either AEH QV-H (— –) or CocE (••••••). A control with no enzyme is also shown (———). All reactions occured in 100 mM PO_4, 20 °C, and were initially at pH 7.0.

While all three reaction mixtures began at the same neutral pH value, they each exhibited different behaviors in the colorimetric assay. The non-enzymatic control showed no appreciable change in pH, which was consistent with the lack of any appreciable conversion. Both the AEH- and CocE-catalyzed reactions initially dropped in absorbance. This decrease in the measured absorbance of phenol red was attributed to a decrease in the pH value of the solution, and was consistent with the carboxylic acid production associated with the hydrolysis reactions (Figure 35). In the AEH-catalyzed reaction, this decrease continued unabated until all activity stopped, which was consistent with reports that the

QV-H variant used did not exhibit measurable *n*-phenylgycyl polymer synthesis activity (Figure 35). [175] For the CocE-catalyzed reaction, the decrease in absorbance was more gradual, and was eventually overtaken by a greater magnitude increase in absorbance after 45 minutes (Figure 40). These results were consistent with the CocE reaction exhibiting competing pH effects, likely due to the presence of both hydrolysis and synthesis reactions (Figure 35). Although the mechanism behind the initial decrease in pH prior to the increase in pH warrants future investigations, it is likely that the rate of primary hydrolysis initially outpaced synthesis due to uncoupling; as excess D-PGME was eliminated, fewer primary hydrolysis side-reactions occurred leading to more efficient catalysis of *n*-phenylgycyl polymer synthesis from the available D-PGME remaining in solution. Synthesis of dipeptide methyl ester by CocE was confirmed by liquid chromatography mass spectrometry (LCMS) analysis (Figure 68 in APPENDIX C, section C.1.6, LC-MS Evidence for Peptide Synthesis).

5.2.5.4 Prediction of Assay Response Curves

The mathematical model described above was used to predict the assay response to hydrolysis of methyl picolinate (Figure 38) under different conditions. Reaction mixtures with concentrations of 20, 50, or 85 mM methyl picolinate in 100 mM sodium phosphate were predicted by modeling and confirmed by experimental measurements of physical samples simulating various conversion levels, prepared by mixing methyl picolinate, picolinic acid, and methanol in predefined ratios as described in the Methods section. The model predictions and experimental measurements of simulated samples are shown in Figure 41.

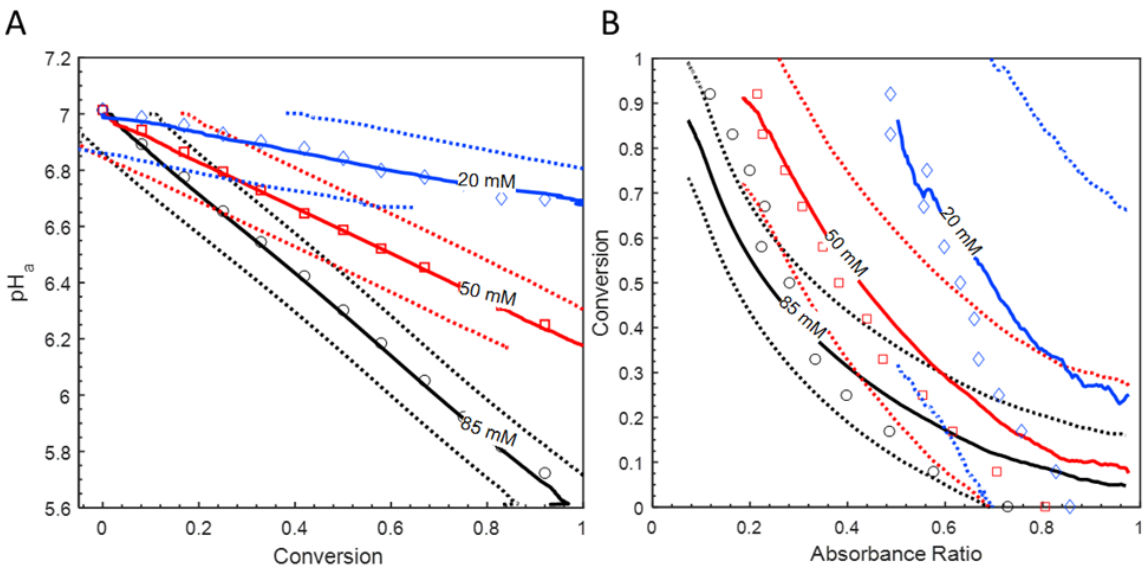


Figure 41: Model predictions and simulated reaction data for the hydrolysis of methyl picolinate. Both plots show model predictions and simulated (non-enzymatic) reaction data for the hydrolysis of either 20 mM (◇), 50 mM (□), or 85 mM (○) methyl picolinate in 100 mM sodium phosphate buffer. Panel A: pH_a vs. Conversion. This is the expected change in measured pH over the course of each reaction. Panel B: Conversion vs. Absorbance Ratio. This is the expected relationship between the reaction conversion and the measured absorbance ratio (557 nm/479 nm) of phenol red in the assay. In both panels: simulated experimental data (manually prepared samples designed to simulate different conversion points in the hydrolysis reactions) are plotted as open symbols, and model predictions are plotted as curves. Solid lines show the mean prediction value, while uncertainty estimates at ±1 standard deviation for 10,000 Monte Carlo samples are indicated by the dashed lines.

In all three predictions (i.e., for initial substrate concentrations of 20, 50, and 85 mM), samples initially at neutral pH begin to acidify as hydrolysis proceeds, and the model predictions match the experimentally measured samples within the estimated prediction error. In general, the magnitude of the pH change is an effect of the background buffer capacity as well as the pK_a values and concentrations of the hydrolysis reactants and products. For a given (nonzero) conversion level, samples prepared with more

substrate will have a greater amount of carboxylic acid product, and thus a greater propensity to overcome the pH-stabilizing effect of the phosphate buffer. Both the predictions and experimental data support this reasoning, and it is apparent in Panel A of Figure 41 that the 85 mM curve has the greatest magnitude slope, and 20 mM the least. The range of the assay within the pH-range of the colorimetric indicator (here, phenol red) dictates the shape of the plot of absorbance ratio vs. reaction conversion. As the pH response in Panel A is relatively linear, and the absorbance response of the indicator is sigmoidal (Figure 62 and Figure 63 in APPENDIX C, section C.1.4, Validation of Phenol Red), the absorbance vs. conversion relationships for these samples are distorted and transformed sigmoidal sections. The nonlinear and widely variable nature of the relationship between the observable color change and the reaction progress is the reason that either experimental or simulated standard curves, or the pH equilibrium model described in this work are required to reliably interpret experimental assay results.

The slope of the relationship between pH and conversion is inherently linked to the sensitivity of the assay, with greater magnitude changes in pH corresponding to a more sensitive assay. This result is captured by the width of the uncertainty bands in Panel B of Figure 41, which are widest for the sample with 20 mM substrate, and narrowest for the sample with 85 mM substrate.

5.2.5.5 Hypothesis Testing and Interpretation of Assay Results

Two example assay samples were prepared to evaluate whether CocE would be classified as a ‘hit’ or a ‘miss’ in a first-pass high throughput screen (evaluating either

enzyme variants or substrate alternatives). Reaction mixes containing phenol red, 50 mM methyl picolinate, and either CocE or a non-enzymatic control in 100 mM sodium phosphate were tracked in a plate reader over several hours. The ratio of optical absorbance (557 nm)/(479 nm) is shown over time in the top panel of Figure 42. The non-enzymatic sample maintains a relatively steady absorbance ratio while the absorbance ratio of the sample containing CocE drops over several hours. Using the model predictions of conversion against absorbance ratio (Figure 41, Panel B) and the supplied spreadsheet for data interpretation and hypothesis testing as described in the Methods section, these absorbance ratios were converted to predictions of conversion with associated levels of prediction uncertainty (Figure 42, middle panel). The results of automatic hypothesis testing, comparing the enzymatic and non-enzymatic samples, are shown in the bottom panel of Figure 42, and the null hypothesis is rejected after 20 minutes indicating that CocE has a statistically significant influence on reaction conversion, and is therefore a 'hit' in this assay. This result is consistent with the established function of CocE as a catalyst for the hydrolysis of methyl picolinate, and supports the use of pH-based colorimetry with model predictions and hypothesis testing for enzyme or substrate screening.

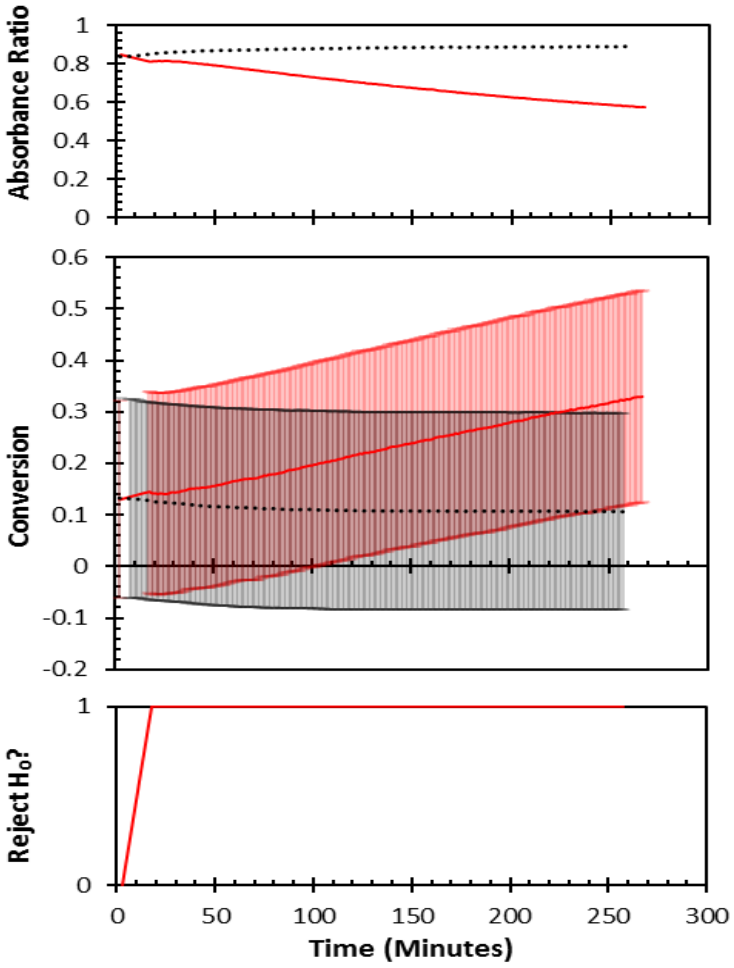


Figure 42: Use of model predictions to perform hypothesis testing. The top panel shows the absorbance ratio (557 nm/479 nm) over time for the hydrolysis of 50 mM methyl picolinate in 100 mM sodium phosphate by CocE (solid red), and non-enzymatic control (dotted black). In the center panel, these data are transformed into hydrolysis conversion, with error bars at ± 1 standard deviation, after Monte Carlo simulation. The bottom panel illustrates how Welch's t-test rejects the null hypothesis within twenty minutes using a p-value threshold of $\alpha = 0.05$.

5.2.6 Discussion

5.2.6.1 Summary of Results

The experimental results presented provide general support and validation for the application of pH-based colorimetry and modeling to high throughput reaction screening. Several intermediate steps were taken to build up to the use of pH and phenol red to track enzymatic reactions, as well as to model and verify predictions of assay behavior for use in high throughput hypothesis testing. Multiple enzymatic hydrolysis and synthesis reactions, listed in Table 14, were used to validate each of the different parts of this work. Phenol red was demonstrated to predictably reveal – but not interfere with – enzymatic synthesis and hydrolysis reactions (reactions 1 and 2 in Table 14). Colorimetry and HPLC were shown to produce comparable results for kinetic analysis of both CocE and AEH (reactions 2 and 3). Colorimetry was also shown to differentiate between reaction regimes dominated by hydrolysis or synthesis activity (reactions 4 and 5).

Table 14: Hydrolysis and synthesis reactions evaluated in this study.

No.	Substrate	Enzyme(s)	Expected Products	Reaction Type
1	D-phenylglycine methyl ester & 6-Aminopenicillinate	AEH	Ampicillin & Methanol	Synthesis
2	Methyl picolinate	CocE	Picolinate & Methanol	Hydrolysis
3	Ampicillin	CocE, AEH	6-aminopenicillinate & D-phenylglycine	Hydrolysis
3	D-phenylglycine methyl ester	AEH	D-phenylglycine & Methanol	Hydrolysis
5	D-phenylglycine methyl ester	CocE	Di- and tri-peptides & Methanol	Synthesis

A generalizable mathematical model was shown to accurately predict the response of the colorimetric assay to experimental samples simulating reaction 2 at various levels of conversion and performed under varied conditions. Finally, an automated hypothesis test correctly classified CocE as having activity on methyl picolinate in an assay trial (reaction 2). These data are the result of a limited range of experiments, but the principles behind them are sufficiently general to warrant further investigation into the combination of pH-based colorimetry and modeling of pH equilibria for high throughput assays and screening tests.

5.2.6.2 Limitations of the pH-Based Assay

The framework we present for colorimetric pH-based assays have the potential to accelerate the development of hydrolase enzymes, such as those utilized for the biocatalytic synthesis of β-LA, by enabling efficient high-throughput analysis of enzyme variants, substrates, or reaction conditions. These benefits notwithstanding, there are some general cases in which the assay or mathematical model presented here may not be appropriate. It is inadvisable to use this assay for reactions with unknown side product formation, because the exact hydrolysis stoichiometry and the acidity constants of all ionizable species are necessary for the prediction of assay response curves. Additionally, the model presented here should not be used for reactions that will exceed ionic strengths of 0.5 M – the recommended limit of Sun's extended Debye-Hückel parameterization. [156] Beyond this range, specific ion interactions may convolute simplified attempts at activity correction,

and more detailed models are recommended. Finally, even when all of the above criteria are met, a particular set of assay conditions might not provide satisfactory sensitivity for a particular reaction. In such cases, slight adjustments to experimental conditions – such as reducing the buffer concentration or starting at a different pH value – may resolve these issues.

In general, the experimental results in Figure 40 indicate that this assay approach would be effective for researchers looking to increase either amide synthesis or ester hydrolysis activity by compressing results into simple increases or decreases in absorbance. Nevertheless, there are limitations that must be considered: the assay would not be able to capture equal and simultaneous increases in both activities due to their opposing effects on solution pH, and certain hydrolysis reactions, such as those of amides, produce equimolar amounts of acid and base, significantly adding to the buffering capacity of the reaction system around a certain pH value. If that particular pH value happens to be the same as the initial pH of the assay, then the pH will not change as a function of conversion. Mathematically speaking, the algebraic solutions to the pH equilibrium model become undefined, as the physical system they are modeling adopts a one-to-many mapping from pH to conversion. This effect is illustrated in Figure 43, which shows the predicted pH response to the hydrolysis of 300 mM acetamide. Acetamide hydrolysis produces equimolar acetic acid and ammonia (Figure 44), which are commonly used together to buffer solutions to pH 7.0. Therefore, no matter the starting pH, the solution will tend towards neutrality. Since CocE is most active at neutral pH, the assay must start at pH 7, meaning that the predicted assay response is undefined.

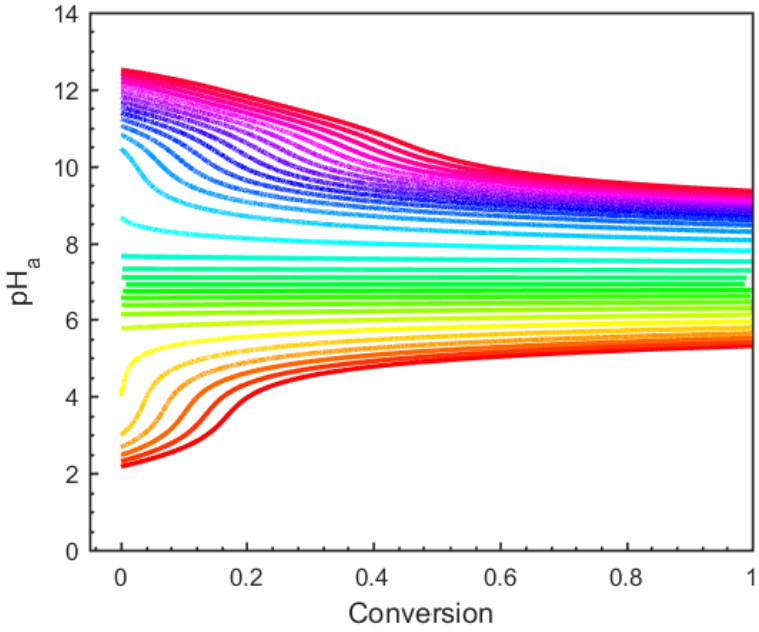


Figure 43: Predicted pH response to the hydrolysis of acetamide. The predicted pH response to conversion for hydrolysis of 300 mM acetamide is shown over a wide (exaggerated, for emphasis) range of starting pH values. The tendency of the pH value towards neutral illustrates a fundamental limitation of the pH-equilibrium model and colorimetric assay.

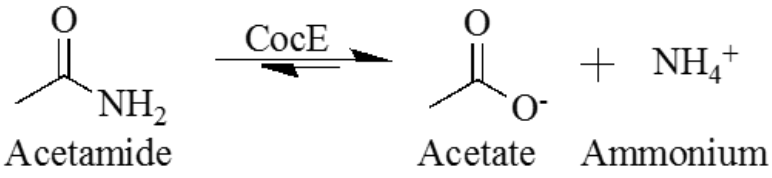


Figure 44: Wild-type CocE can catalyze the hydrolysis of acetamide into acetate and ammonium.

Finally, this approach should only be used for first-pass screening, as the extended Debye-Hückel method used here is an empirical regression of experimental data, and may not provide satisfactory results for all compounds. As with most colorimetric assays, insoluble and precipitating components will hinder correct measurements; assays should be performed within the solubility limits of all reagents.

5.2.6.3 Advantages of the pH-Based Assay over Existing Methods

In traditional HPLC-based assays, the elution time and absorbance level of each compound are affected by the pH of the mobile phase, as well as the column condition and temperature. Additionally, chromatography columns rely on dynamic cross-linked functional groups that may exhibit varied behavior under different conditions of mobile phase, storage solvents, upkeep level, and age. These variables are difficult to control, and failure to do so can result in peak-splitting, degradation of assay quality, and inconsistent results. Furthermore, designing separation methods for and producing standard curves for a wide variety of possible reactants and products is time consuming, and may require switching between multiple HPLC columns or mobile phases. Finally, the challenges of using HPLC for high-throughput analysis are exacerbated by the prolonged sample run-time – typically requiring at least 30 minutes to collect a single data point.

As optical measurements in general are simpler and faster than HPLC, switching a first-pass screening operation from HPLC to colorimetry could result in a significant reduction in time and material costs, as well as the volume of data, in comparison to HPLC. Using the phenol red assay, multiple enzymatic reactions can be loaded into a standard

well-plate and can be reliably qualified as ‘hit’ or ‘miss’ in 100 minutes or less. Additionally, if hypothesis testing is performed in real-time during absorbance measurement, then a consistently low p-value may permit early termination of a sample. This is illustrated for the hydrolysis of 50 mM methyl picolinate by CocE over several hours in the bottom panel of Figure 42. In this example, it is clear after less than an hour that the hydrolysis of 50 mM methyl picolinate should be considered a positive ‘hit,’ and that the assay (or monitoring of that particular set of samples) may be terminated early. There are also other benefits to continuously monitoring assay samples: a reduced number of samples can be used, as repeated observation of an individual sample provides an effective denoising, and time-series data provide opportunities for other uncertainty-management approaches such as low-pass filters, or more sophisticated statistical methods involving sequential analysis.

Alternative high throughput colorimetric assays exist for these and other reactions, but the current paradigm in which every reaction must have its own specialized assay may not be the most efficient use of resources. The method presented here relies only on acid-base equilibria, and is therefore generalizable to the analysis of many reactions and enzymes. The mathematical model can predict direct mappings between optical absorbance and conversion for a variety of reactions without necessarily requiring experimental standards, and cases in which the colorimetric assay will not work can frequently be determined in advance of any experimentation in the laboratory.

5.2.7 Conclusions

The use of phenol red as an indicator for high-throughput pH-based colorimetric substrate scanning and activity screening of enzyme variants was demonstrated to be successful. This assay can be employed to detect reactions catalyzed by AEH and CocE as well as other enzymes within the α/β hydrolase family, and can be used to distinguish hydrolysis from synthesis activity. A thermodynamic pH equilibrium and reaction-conversion model compresses the multitude of variables embodied in a reaction system into simple and direct relationships between optical absorbance and reaction conversion. This model provides insights into assay behavior that facilitate planning of experiments. Model predictions can be used in lieu of experimental standard curves – the collection of which might detract from the otherwise high-throughput nature of the assay – and facilitate real-time hypothesis testing to classify samples as hits or misses. The approach used here is sufficiently general to enable use of colorimetric pH-based assays for rapid screening of other aqueous reactions, which result in pH changes.

5.3 Calculation of Ionic Equilibria and Determination of Acid Dissociation Constants by Isothermal Titration Calorimetry

5.3.1 Statement of Authorship

The text in this section was coauthored by the following: Harrison Bellow Rose, [a] Madison M. Wilber, [a] and Andreas S. Bommarius. [a, b]

[a] School of Chemical & Biomolecular Engineering Georgia Institute of Technology, Atlanta, Georgia 30332, United States. [b] School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States.

Experimental work and analysis was performed by Harrison B. Rose, with assistance from undergraduate research assistant Madison M. Wilber.

5.3.2 Abstract

A model is presented representing the equilibrium pH of an aqueous solution as a sparse eigenvalue problem. This calculation is applied to model the enthalpies of mixing and diluting aqueous buffers. The addition of a transient mass balance captures the behavior of isothermal titration calorimetry (ITC) experiments in which different buffers are present in the titration syringe and sample cell. This model and ITC technique are illustrated with an experiment to determine the acidity constants of three carboxylic acids in the furan series: 2-furancarboxylic acid (FA), 5-formyl-2-furancarboxylic acid (FFA), and 2,5-furandicarboxylic acid (FDCA). Acidity constants were also evaluated by a spectrophotometric method, and the thermodynamic pKa values of the analytes were found

to be 3.1 for FA, 2.2 for FFA, and 2.1 and 3.4 for the first and second ionization steps of FDCA.

5.3.3 Introduction

Acid-base equilibria are ubiquitous in aqueous environments, and a strong fundamental understanding of their behavior is requisite to advancing knowledge and application development in fields ranging from limnology and climatology to biopharmaceutical formulation.[190-191] Although ionic equilibria abound in naturally occurring and engineered systems, only limited experimental and computational tools exist with which to probe and design them; they are under-leveraged in many applications.

Tabulated thermodynamic data provide a good illustration of this deficiency. Heats of dilution of aqueous electrolytes, including strong and weak acids and simple salts, are typically reported in the literature as *apparent* relative molal values.[111, 192-195] A lack of robust methods for calculating ionic equilibria makes it difficult to deconvolute the heats of ionization and neutralization (here collectively called 'pH effects') from the excess enthalpies of mixing. Although some studies deliberately separate these effects[196-198] or give brief mention,[199] others – even those with similar systems – neglect them altogether.[200-201] Their consideration still has not been widely adopted and incorporated into common sources of thermodynamic data today, unnecessarily restricting the ways in which these data can be applied to other problems.

Looking beyond the *apparent* effects of dilution could impact other areas of research, such as the biological sciences. Isothermal titration calorimetry (ITC), in which

a titrant is repeatedly injected into a sample cell, is a powerful tool for evaluating both binary solution thermodynamics [202] and binding and interaction properties of biomolecules. [203] However, its use in most investigations is constrained by the requirement that titrant and titrand are both comprised of the same bulk buffer (pH, composition, ionic strength) because there is not an accepted method of accounting for the complex pH effects that would otherwise occur upon mixing. Investigations involving labile protons are thus typically limited to proton linkages associated with protein-ligand binding, [204] although there are some exceptions which are discussed below. While this restriction tends to produce very elegant ligand binding experiments, it limits overall breadth of experiments that can be reasonably conducted using ITC. For example, binding studies between pharmaceutical candidates and protein targets may be limited by poor solubility of the ligand in the protein buffer. By enabling the use of an alternate, solubility-enhancing, buffer in the ITC syringe, higher concentrations of the drug candidate might be evaluated, or experiments could be conducted under conditions more closely reflective of possible drug product formulations.

There are a few notable examples in the literature related to determination of pK_a values by ITC, but they are all restricted in design by ionic background. Tajc, et al. (2004) [205] evaluated the pK_a of the thiol side chain of cysteine residues by measuring the heat of thioether formation when titrating iodoacetamide into samples of cysteine prepared in buffers over a range of pH values. The analysis performed in that study assumes that the pH value of each sample does not shift over the course of multiple titrations, and neglects enthalpies associated with interactions between solutes. These may be reasonable

assumptions, but it is difficult to prove what conditions must be met for that to be the case, or the degree to which the results and conclusions would be impacted if it were not. Both Shoghi, et al. (2009) [206] and Zheng, et al. (2015) [207] later determined pK_a values of weak acids using more traditional single-site-binding experiments by titrating hydrochloric acid directly into aqueous solutions or methanol/water mixtures containing the conjugate base of each analyte. In contrast to the previous example, these experiments require that the pH value within the calorimeter cell shifts with each injection of strong acid, but non-reactive interaction enthalpies are neglected, and no other solutes can be present.

To address the shortcomings in the handling of ionic equilibria, we present a new mathematical approach for the evaluation of multicomponent aqueous buffer systems, including heats of mixing and pH effects in batch systems or in unsteady continuously stirred tank reactors (CSTR, as an ITC could ostensibly be described). This framework was demonstrated by application to a new method for determining pK_a values and neutralization enthalpies of three poorly soluble weak acids in the furan series using ITC: 2-furancarboxylic acid (FA), 5-formyl-2-furancarboxylic acid (FFA), and 2,5-furandicarboxylic acid (FDCA). In the new method presented here, weakly buffered solutions with different solutes can be present in both the titration syringe and the calorimeter cell, and heats of dilution and neutralization of all components are accounted for simultaneously. pK_a values of the furancarboxylic acids are left as adjustable parameters in the model, and the fitted values for FA and FFA were shown to be consistent with those obtained by spectrophotometry.

5.3.4 Methods

5.3.4.1 Materials

ACS grade anhydrous citric acid, enzyme grade potassium chloride and potassium hydroxide, and LC/MS grade water were acquired from Fisher Chemical (Fair Lawn, NJ). Minimum 98% Sodium hydroxide pellets, minimum 98% 2-furoic acid, and analytical 0.1 M hydrochloric acid were acquired from Sigma-Aldrich (St. Louis, MO). 5-formyl-2-furancarboxylic acid and 2,5-furandicarboxylic acid were both minimum 98%, and were purchased from TCI America (Portland, OR). ACS-grade sodium chloride was purchased from BDH VWR Analytical (Radnor, PA). Chemicals were used as-received; no additional preparatory purifications or analyses of purity were performed.

5.3.4.2 Preparation of Titrant and Titrand Solutions

In a more traditional pH titration, the analyte acid would reside in the titrand and the titrant would consist of a strong acid or base. In this work, however, the same analyte solution is titrated into multiple buffers, one at a time. The titrant solutions consisted of 2.0 mM stocks of FA, FFA, or FDCA. Titrants were prepared in LC/MS grade water using an analytical balance and calibrated glassware. The titrand solutions were 31 buffers that were prepared from citric acid, sodium hydroxide, hydrochloric acid, and sodium chloride. All buffers were prepared with a final citric acid concentration of 20.0 mM. Sodium hydroxide and hydrochloric acid concentrations were selected to achieve a range of pH values from 1 to 5.5, which was anticipated to encompass the pK_a values of the analyte acids. Sodium chloride was added to bring the total ionic strength of each buffer to 100 mM. Conditions

were calculated according to a reduced version of the equations discussed in the modeling section.

5.3.4.3 Potentiometric Electrode Calibration

A glass Ag/AgCl electrode was calibrated against proton concentration (rather than activity) using an autotitration apparatus that was provided by the Fahrni Lab (School of Chemistry & Biochemistry, Georgia Institute of Technology).

0.10 M solutions of KOH and KCl were prepared in calibrated glassware. The KOH solution was sparged with argon for 10 minutes to reduce carbonate content. Each solution was loaded into a Brinkmann Metrohm 665 liquid-handling dosimat. 20.0 mL of 0.10 M KCl and 1.000 mL of 0.10 M HCl were loaded into a water-jacketed titration vessel that was maintained at 25 °C, flushed with argon, and stirred with a Teflon PTFE-coated magnetic stirrer. Electrode potential was recorded as 40.0 μL additions of KOH were made with the dosimat to a final added titrant volume of 2.000 mL. This process was repeated three times.

The software package, GLEE3, developed by Gans and O'Sullivan (2000) [187] can analyze data from a single standardizing titration. The approach outlined in their paper was adapted to fit triplicate calibration titrations simultaneously for this study. Following the convention established in GLEE3, data outside of the pH ranges 2.5-4.0 and 10.7-11.3 were excluded from the standardization process.

5.3.4.4 Determination of Proton Concentration

The same electrode and thermostatic vessel used for calibration were used to measure voltage potentials for each of the 31 buffer solutions at 25 °C. Proton concentrations (indicated by the base-10 logarithm as pH_c) were calculated from the calibration curve.

5.3.4.5 Spectrophotometric Determination of pK_a Values

pK_a values for FA, FFA, and FDCA were identified by population balance and halochromic shift using a Cary 8454 UV-Vis spectrophotometer with a photodiode array (PDA) detector (Agilent Technologies, Santa Clara, CA) and a quartz cuvette with a path length of 1 cm. 31 sets of 4 samples each were prepared by aliquoting 40 µL of either water or one of the three analyte stock solutions into 1.6 mL of each of the 31 buffer solutions. Samples were mixed thoroughly and incubated at 25 °C. The instrument was blanked with water, and the full absorbance spectrum of each sample was measured. Spectra obtained from samples without analyte (i.e., those containing only diluted buffer) were subtracted from those samples of the corresponding buffer containing dilute analyte.

5.3.4.6 Isothermal Titration Calorimetry Measurements

A MICROCAL PEAQ-ITC (Malvern Instruments Inc., Westborough, MA) power-compensation-type isothermal titration calorimeter with a cell volume of 200 µL was used for all calorimetric experiments. Each experimental measurement entailed four distinct and sequential 1 µL titrant injections into the calorimeter cell, which was loaded with titrand

solution and maintained at 25 °C. The instrument was set with a stir speed of 750 rpm, and injections were 1 second in duration, and proceeded every 120 seconds with an initial delay of 120 seconds. Each consisted of four injections, and data from the first injection was discarded, in accordance with standard convention. [205] Each experiment was preceded by a thorough cleaning of the instrument: the cell and syringe were washed using the default wash cycle and an additional rinse cycle, followed by three manual flushes of the cell and syringe with their respective working fluid for the upcoming experiment. The reference cell was rinsed and filled with LCMS-grade water prior to use. All solutions were incubated at 25 °C and vacuum-degassed for 5 minutes immediately prior to use using a MICROCAL THERMOVAC (Malvern Instruments, Inc.). Control experiments were performed to determine the frictional heat of injection (three injections of water into water) and the heat of diluting each of the buffers (water was injected into each buffer). Instrument-reported values in microcalories were converted to joules using a factor of $4.184 \times 10^{-6}\,\text{Joules}/\mu\text{cal}$.

5.3.5 Mathematical Considerations

5.3.5.1 Analysis of Spectrophotometric Data

For a monoprotic weak acid in solution with concentration-based $pH_{c,i}$, the distribution fractions of acid in the protonated and deprotonated states, ${}^{HA}f^{pH_{c,i}}$ and ${}^{A^-}f^{pH_{c,i}}$ can be described by Equations (31) and (32):

$$ {}^{HA}f^{pH_{c,i}} = \frac{[HA]}{[HA]+[A^-]} = \frac{10^{-pH_{c,i}}}{10^{-pH_{c,i}} + 10^{-pK_a^c}} \tag{31} $$

$$ {}^{A^-}f^{pH_{c,i}} = \frac{[A^-]}{[HA]+[A^-]} = \frac{10^{-pK_a^c}}{10^{-pH_{c,i}} + 10^{-pK_a^c}} \tag{32} $$

where pK_a^c is the base 10 logarithm of the acid association constant calculated on a concentration basis. Similarly, the distribution fractions of a diprotic weak acid can be expressed by Equations (33)-(35), where $pK_{a,1}^c$ and $pK_{a,2}^c$ are the logarithms of the first and second concentration-based acid association constants, respectively.

$$ {}^{H_2A}f^{pH_{c,i}} = \frac{10^{-2\times pH_{c,i}}}{10^{-pK_{a,1}^c - pK_{a,2}^c} + 10^{-pK_{a,1}^c - pH_{c,i}} + 10^{-2\times pH_{c,i}}} \tag{33} $$

$$ {}^{HA^-}f^{pH_{c,i}} = \frac{10^{-pK_{a,1}^c - pH_{c,i}}}{10^{-pK_{a,1}^c - pK_{a,2}^c} + 10^{-pK_{a,1}^c - pH_{c,i}} + 10^{-2\times pH_{c,i}}} \tag{34} $$

$$ {}^{A^{2-}}f^{pH_{c,i}} = \frac{10^{-pK_{a,1}^c - pK_{a,2}^c}}{10^{-pK_{a,1}^c - pK_{a,2}^c} + 10^{-pK_{a,1}^c - pH_{c,i}} + 10^{-2\times pH_{c,i}}} \tag{35} $$

In a cell of path length ℓ, the total absorbance $Abs_{\lambda_j}^{pH_{c,i}}$ at the i^{th} pH_c value, $pH_{c,i}$, and the j^{th} wavelength, λ_j, can be expressed as a linear combination of the molar absorptivities, ε, of the protonated and deprotonated states. This relationship is provided in Equation (36), where S is used as a shorthand for the set of accessible protonation states (i.e, $\{HA, A^-\}$ or $\{H_2A, HA^-, A^{2-}\}$ for a mono or diacid, respectively).

$$\frac{Abs_{\lambda_j}^{pH_{c,i}}}{[A]_{tot}\ell} = \sum_{s} {}^{S}f^{pH_{c,i}} \times {}^{S}\varepsilon_{\lambda_j}, \qquad \forall S \in \begin{cases} \{HA, A^-\}\ if\ \ monoacid \\ \{H_2A, HA^-, A^{2-}\}\ if\ \ \ diacid \end{cases} \tag{36}$$

The concept of Equation (36) is not new to this work, and has an extensive history of use for the characterization of acids, though it is typically limited to a single wavelength and data are processed through algebraic transformations for linear regression.[208] However, with PDA detectors that provide entire spectra with a single measurement and matrix algebra toolkits such as MATLAB, it may be more reasonable nowadays to analyze hundreds or thousands of data points at once than it would have been before these tools were available.

Given a guess of pK_a^c for a monoacid and the experimental absorbance and pH_c values for each buffer, Equations (31), (32), and (36) can be used to obtain the exact least squares solution for the molar absorptivities of each protonation state at all wavelengths simultaneously. By alternately guessing values for pK_a^c, and recalculating the molar absorption spectra, a nonlinear solver can be used to identify the pK_a^c guess that minimizes the residual error between the experimental spectra and those predicted by Equations (31),

(32), and (36). Similarly, the two pK_a^c values for a diacid can be fitted by providing two guesses and using Equations (33)-(35) in lieu of (31) and (32).

Once the optimal value (or values) of pK_a^c is (are) found and corresponding values of ${}^S\varepsilon_{\lambda_j}$ calculated for all protonation states and wavelengths, Equations (36) can also be used to calculate the exact least-squares values of ${}^S f^{pH_{c,i}}$ for all protonation states at all values of $pH_{c,i}$, deconvoluting the acid species distribution that was present in each sample.

The resultant values for pK_a^c were extrapolated to a thermodynamic basis, pK_a^{th}, using the guidelines presented in CHAPTER 5, section 5.2.4.7, Modeling and Prediction of Assay Response Curves and APPENDIX C, section C.2.2, Standardization of pK_a Values from the Literature.

5.3.5.2 Analysis of Calorimetric Data

ITC typically reports the heating rate, Q, necessary to maintain the temperature of the cell at the setpoint temperature. This heating rate thus depends on a multitude of factors, including the rate of heat production within the cell, the conductivity of the cell and fluid, and the instrument set parameters. By integrating the heat of each injection, any rate-dependent features of the data can be neglected, and each injection can be treated as a discrete event with an initial equilibrium state and a final equilibrium state. Since enthalpy is a state function, the order and rate of events within an injection are not important.

Several heat effects contribute to ITC data including: (A) the temperature difference between the injected fluid and the cell; (B) the frictional heats of injection and stirring; (C)

the excess enthalpies of mixing and dilution of the solutes within the titrant and titrand, and (D) the heats of reactions that occur within the cell, here limited to ionization and neutralization. Any model that purports to reproduce ITC data with mixed buffers should be able to capture each of these effects, such that their sum reproduces the experimentally determined integrated heats of injection. Assuming a constant room temperature and neglecting differences in buffer viscosities, the combined effects of (A) and (B) can be captured by control experiments with water in both the syringe and the sample cell, and the integrated heat of injection can then be subtracted from the integrated heats collected during other experiments. Determination of (C) and (D) are more involved and require accounting for the concentrations of, and interactions between, species in the titrant and titrand.

5.3.5.3 Calculation of Sample Analytical Composition After Injection

The calorimeter used in this experiment is initialized with a full sample cell, and each injection displaces an equal amount of fluid into an overflow trough such that total sensible volume, V_{cell}, remains constant, as shown in Figure 45.

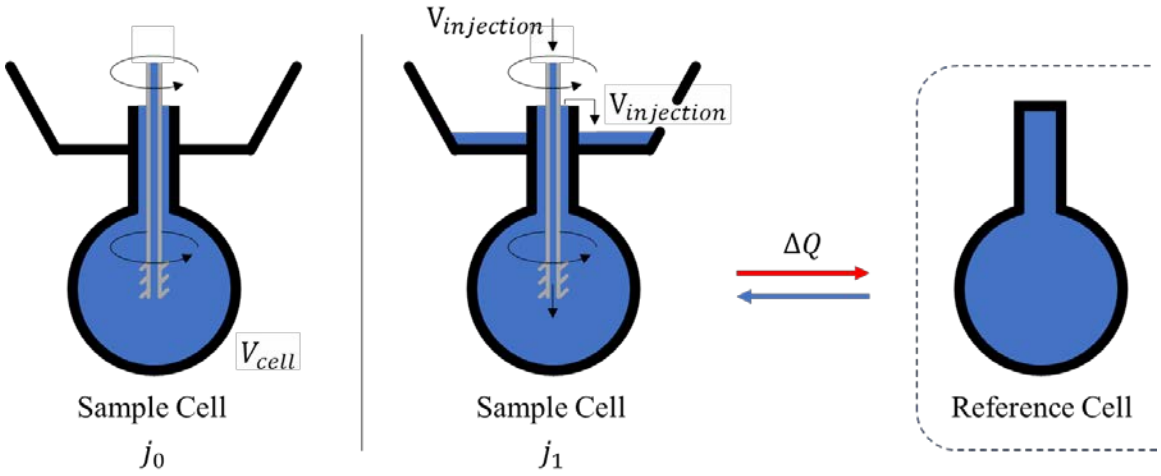


Figure 45: Cartoon schematic of an isothermal titration calorimeter before and after injection. The isothermal titration calorimeter has a fixed volume V_{cell}. The fluid added from the initial state j_0 to the first injection j_1 has a volume of $V_{injection}$, and an equal volume is displaced into the overflow trough. The heat of the injection, ΔQ, is measured as the heating or cooling duty relative to the reference cell. The needle of the injection syringe has flutes and spins to ensure that the sample cell is well mixed.

Conceptually, then, the mass balance equation of a transient CSTR could be used to represent a steady string of injections, and the change in the concentration of a species i in the cell, $C_{i,cell}$, resulting from a single titrant addition of volume $V_{injection}$ can be approximated by applying a finite difference approximation with a time step that encompasses the injection, as shown in Equation (37), where j is the injection number and $C_{i,syringe}$ is the concentration of species i in the titrant:

$$C_{i,cell}(j+1) = \left(\frac{V_{injection}}{V_{cell}}\right) C_{i,syringe} + \left(1 - \frac{V_{injection}}{V_{cell}}\right) C_{i,cell}(j) \qquad (37)$$

The heats of dilution (C) and those of ionization and neutralization reactions (D) can be accounted for in a sequence of separate steps. First, the equilibrium pH and distributions of buffer species are calculated for the titrant in the syringe and titrand in the sample cell prior to injection. An “injection” is made, and the new composition in the sample cell is calculated according to Equation (37), absent any consideration of steps toward reaching an ionic equilibrium. The equilibrium composition prior to injection and the unequilibrated composition after injection are then used to calculate the heats of dilution (C) of each solute. Binary interaction parameters are added to capture excess heats of mixing, and the resulting enthalpies are summed. Next, the final equilibrium pH and buffer species distributions are calculated and used to determine the net extent of reaction necessary for each buffer ionization step to bring the unequilibrated system to the calculated equilibrium. Finally, the heats of ionization are multiplied by the extent of the corresponding reaction to yield component (D), and all resulting enthalpies are summed to a value that should equal the experimentally determined integrated heat of injection.

5.3.5.4 Calculation of Ionic Equilibrium

Calculating the equilibrium pH of an aqueous mixture of buffer species is a surprisingly challenging problem. The Henderson-Hasselbalch equation is frequently used to estimate the equilibrium position of monoprotic buffers, but it relies on the implicit assumptions that dissociation and hydrolysis are negligible – assumptions that break down and lead to inaccurate calculations even under routine conditions. [188, 209] As others have noted before, [188, 209-210] more precise calculations are possible with complete accounting of all (de)protonation steps and a charge balance. For a given buffer solution, this entails one

equation for the self-ionization of water, one for each ionization step of the buffer components, an analytical total of the buffer concentration, and a charge balance which may include co-ions in solution such as sodium or chloride. The equations that must be simultaneously solved for this system are described below:

$$[B]_{tot} = [H_3B] + [H_2B^-] + [HB^{2-}] + [B^{3-}] \quad (38)$$

$$K_{1,B}^{th} = \frac{[H^+][H_2B^-]}{[H_3B]} K_{1,B}^{\gamma} \quad (39)$$

$$K_{2,B}^{th} = \frac{[H^+][HB^{2-}]}{[H_2B^-]} K_{2,B}^{\gamma} \quad (40)$$

$$K_{3,B}^{th} = \frac{[H^+][B^{3-}]}{[HB^{2-}]} K_{2,B}^{\gamma} \quad (41)$$

Where $[H^+]$ is hydronium ion concentration; $[B]_{tot}$ is the analytical concentration of citric acid buffer; $[H_3B]$, $[H_2B^-]$, $[HB^{2-}]$, and $[B^{3-}]$ are the four ionization states of citric acid corresponding to the thermodynamic acid dissociation constants $K_{1,B}^{th}$, $K_{2,B}^{th}$, and $K_{3,B}^{th}$; and $K_{1,B}^{\gamma}$, $K_{2,B}^{\gamma}$, and $K_{3,B}^{\gamma}$ are ratios of activity coefficients as described in CHAPTER 5, section 5.2.4.7, Modeling and Prediction of Assay Response Curves and APPENDIX C, section C.2.2, Standardization of pK_a Values from the Literature. The thermodynamic dissociation constants for citric acid were determined from literature values according to the relation $K_{i,B}^{th} = 10^{-pK_{i,B}^{th}}$, where $pK_{i,B}^{th}$ are 2.859, 3.880, and 5.160 for the $i = 1^{st}$, 2^{nd}, and 3^{rd} deprotonation steps of citric acid, respectively. [211]

The distribution of analyte acid (here, any of the furancarboxylic acids) must be similarly accounted for:

$$[A]_{tot} = [H_2A] + [HA^-] + [A^{2-}] \tag{42}$$

$$K_{1,A}^{th} = \frac{[H^+][HA^-]}{[H_2A]} K_{1,A}^{\gamma} \tag{43}$$

$$K_{2,A}^{th} = \frac{[H^+][A^{2-}]}{[HA^-]} K_{2,A}^{\gamma} \tag{44}$$

Where $[A]_{tot}$ is the analytical concentration of analyte; $[H_2A]$, $[HA^-]$, and $[A^{2-}]$ are the ionization states of the analyte, and $K_{1,A}^{th}$ and $K_{2,A}^{th}$ and $K_{1,A}^{\gamma}$ and $K_{2,A}^{\gamma}$ are equilibrium constants and lumped activity coefficients as before. If the analyte has only one labile proton (as is the case for FA and FFA), the second acid dissociation constant can be set to any arbitrarily low value corresponding to an unreachable pH, such as 10^{-20}.

The self-ionization of water must also be included:

$$K_w^{th} = [H^+][OH^-]K_w^{\gamma} \tag{45}$$

Where K_w^{th} is the thermodynamic ionization constant for water, $[OH^-]$ is the hydroxide ion concentration, and K_w^{γ} is an activity correction.

And finally, a charge balance constraint prevents the system from accumulating a net charge:

$$[H^+] + [Na^+] = [OH^-] + [Cl^-] + [H_2B^-] + 2[HB^{2-}] + 3[B^{3-}] + [A^-] + 2[A^{2-}] \quad (46)$$

For simple buffers (i.e., a single monoprotic acid in water), analytical expressions for species distribution and equilibrium pH can be found manually by algebraic substitution. [188, 209] However, that approach quickly becomes infeasible for more complex systems, as illustrated by the model discussed in CHAPTER 5, section 5.2.4.7, Modeling and Prediction of Assay Response Curves and the resultant expressions in APPENDIX C, section C.2.6, Mathematica Output: Expressions for Absorbance, Conversion, and Ionic Strength. Incorporation of activity coefficients (i.e., $K_x^\gamma \neq 1$) as performed here can improve model accuracy, but the additional step of revising equilibrium constants with changes in ionic strength necessitates switching to a solver capable of fixed-point iteration. An example of this approach has been demonstrated for calculating the distribution diagram of citric acid in solution, [212] but iterative numerical solutions quickly become unwieldy for larger and larger systems, and specialized solvers are required to accommodate the scaling problems that arise. [210] We have previously demonstrated (CHAPTER 5, section 5.2.4.7, Modeling and Prediction of Assay Response Curves) an approach to evaluating equilibrium pH in multiply buffered systems with hydrolysis reactions using computer-generated analytical expressions, although the expressions are multiple pages long and are not human-readable or particularly elegant (APPENDIX C, section C.2.6, Mathematica Output: Expressions for Absorbance, Conversion, and Ionic Strength).

In the fixed point iterator for citric acid calculations, Apelblat (2014) [212] makes a useful substitution of distribution fractions in place of conjugate base concentrations, e.g. β for ionized states of citric acid buffer and α for ionized states of analyte acid:

$$[H_2B^-] = [B]_{tot}\,\beta_1 \tag{47}$$

$$[HB^{2-}] = [B]_{tot}\,\beta_2 \tag{48}$$

$$[B^{3-}] = [B]_{tot}\,\beta_3 \tag{49}$$

$$[HA^-] = [A]_{tot}\,\alpha_1 \tag{50}$$

$$[A^{2-}] = [A]_{tot}\,\alpha_2 \tag{51}$$

The powerful part of this substitution that we believe has not previously been exploited to full advantage is that it enables the linearization of the entire system of Equations (38)-(46) with respect to proton concentration:

$$(1 - \beta_1 - \beta_2 - \beta_3)\,K_{1,B}^{th}/K_{1,B}^{\gamma} = [H^+]\beta_1 \tag{52}$$

$$\beta_1\,K_{2,B}^{th}/K_{2,B}^{\gamma} = [H^+]\beta_2 \tag{53}$$

$$\beta_2\,K_{3,B}^{th}/K_{3,B}^{\gamma} = [H^+]\beta_3 \tag{54}$$

$$(1 - \alpha_1 - \alpha_2)\,K_{1,A}^{th}/K_{1,A}^{\gamma} = [H^+]\alpha_1 \tag{55}$$

$$\alpha_1\,K_{2,A}^{th}/K_{2,A}^{\gamma} = [H^+]\alpha_2 \tag{56}$$

$$K_w^{th}/K_w^{\gamma} = [H^+][OH^-] \tag{57}$$

$$[OH^-] + [Cl^-] + [B]_{tot}(\beta_1 + 2\beta_2 + 3\beta_3) + [A]_{tot}(\alpha_1 + 2\alpha_2) - [Na^+] = [H^+] \tag{58}$$

This linearized system of Equations (52)-(58) can then be reformulated as an eigenvalue problem $Ax = \lambda x$, [f] which can be solved much more efficiently by existing algorithms (note that the charge balance is now listed first):

[f] This step and the previous were identified for Apelblat's formulation of the citric acid system by Dr. Satin Gungah (Control and Power Group, Department of Electrical and Electronic Engineering, Imperial College, London, UK) in a personal communication with this author.

$$\overbrace{\begin{bmatrix} [Cl^-]-[Na^+] & [B]_{tot} & 2[B]_{tot} & 3[B]_{tot} & [A]_{tot} & 2[A]_{tot} & 1 \\ K_{1,B}^{th}/K_{1,B}^{\gamma} & -K_{1,B}^{th}/K_{1,B}^{\gamma} & -K_{1,B}^{th}/K_{1,B}^{\gamma} & -K_{1,B}^{th}/K_{1,B}^{\gamma} & 0 & 0 & 0 \\ 0 & K_{2,B}^{th}/K_{2,B}^{\gamma} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{3,B}^{th}/K_{3,B}^{\gamma} & 0 & 0 & 0 & 0 \\ K_{1,A}^{th}/K_{1,A}^{\gamma} & 0 & 0 & 0 & -K_{1,A}^{th}/K_{1,A}^{\gamma} & -K_{1,A}^{th}/K_{1,A}^{\gamma} & 0 \\ 0 & 0 & 0 & 0 & K_{2,A}^{th}/K_{2,A}^{\gamma} & 0 & 0 \\ K_{w}^{th}/K_{w}^{\gamma} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}^{\boldsymbol{A}} \times \overbrace{\begin{bmatrix} 1 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \alpha_1 \\ \alpha_2 \\ [OH^-] \end{bmatrix}}^{\boldsymbol{x}} = \overbrace{[H^+]}^{\boldsymbol{\lambda}} \times \overbrace{\begin{bmatrix} 1 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \alpha_1 \\ \alpha_2 \\ [OH^-] \end{bmatrix}}^{\boldsymbol{x}} \quad (59)$$

It may be a mathematical property of systems such as Equation (59) that only a single eigenvalue λ can be found that is both positive and real. This remains to be rigorously proven or disproven, but no evidence to the contrary has arisen in the over 496 [g] configurations evaluated in this study. Equation (59) is unlike most other eigenvalue problem representations of chemical reaction equilibria, which typically begin as ordinary differential equations, because the positive real eigenvalue λ is defined specifically to be a species concentration (the proton or hydronium ion concentration), rather than a more abstract construction representing the trajectory and stability of a transient reaction system in the phase plane.

Eigenvectors are typically solved within a multiplicative constant, but the first entry of unity enables the eigenvector x to be correctly rescaled such that the concentrations of each ionized state of the buffer and analytes are easily recoverable from the values of β_1, β_2, β_3, α_1, and α_2, given that the analytical species totals $[B]_{tot}$ and $[A]_{tot}$ are known from Equation (37). This approach appears to be scalable to buffer solutions with even more components, and it provides a single immediate equilibrium composition if activity coefficients are set at unity. If activity coefficients are incorporated, as they were in this work, the eigenvalue problem can be used as a single calculation step in a fixed-point

[g] Four injections were calculated for each of the four titrants into each of the 31 buffers (a base case of 496 different evaluations of Equation (55)). Furthermore, K^{γ} values were iteratively updated until ionic strength stabilized for each injection, and the entire system was used by a nonlinear solver during parameter fitting with each iteration producing only one eigenvalue that was both positive and real.

iterator as before, and solutions tend to converge rapidly (again, based on the anecdotal observations from the over 496 runs evaluated in this study).

To account for activity coefficients, we have used an adaptation of the extended Debye-Hückel correction fitted by Sun, et al. (1980) [156] that was derived and demonstrated previously (CHAPTER 5, section 5.2.4.7, Modeling and Prediction of Assay Response Curves and APPENDIX C, section C.2.2, Standardization of pK_a Values from the Literature):

$$K^{\gamma} = 10^{(-1.02\times z)\left(\left(\frac{\sqrt{I}}{1+A\sqrt{I}}\right)+BI\right)+CI} \quad (60)$$

Where A, B, and C are general parameters equal to 1.50, -0.09, and -0.09, respectively.

5.3.5.5 Calculation of Reaction Enthalpies

Six neutralization reactions can occur to allow the sample to settle into ionic equilibrium after each injection. Only one equilibrium is defined, and the extent of each reaction, ξ, can be calculated uniquely using the scheme presented in Table 15, given the initial composition (i.e., immediate post-injection and dilution), which is provided by Equation (4), and the final equilibrium composition given by Equation (59). To calculate extents of reaction, compositions must be represented by the number of moles, $n_{(i)} = C_i V$, where C_i is the concentration of the species in question, and V is the reacting volume. For dilutions within the ITC, V is fixed at V_{cell}.

Table 15: Neutralization reactions and their extents in moles.

No.	Reaction	Explanation	Extent of Reaction
1	$H^+ + OH^- \rightarrow H_2O$	Neutralization of water	$\xi_1 = n_{(H^+)_{initial}} - n_{(H^+)_{final}}$
2	$H_3B + OH^- \rightarrow H_2B^- + H_2O$	First deprotonation of citric acid	$\xi_2 = n_{(BH_3)_{initial}} - n_{(BH_3)_{final}}$
3	$H_2B^- + OH^- \rightarrow HB^{2-} + H_2O$	Second deprotonation of citric acid	$\xi_3 = n_{(BH_2^-)_{initial}} - n_{(BH_2^-)_{final}} + \xi_2$
4	$HB^{2-} + OH^- \rightarrow B^{3-} + H_2O$	Third deprotonation of citric acid	$\xi_4 = n_{(BH^{2-})_{initial}} - n_{(BH^{2-})} + \xi_3$
5	$H_2A + OH^- \rightarrow HA^- + H_2O$	First deprotonation of analyte acid	$\xi_5 = n_{(AH_2)_{initial}} - n_{(AH_2)_{final}}$
6	$HA^- + OH^- \rightarrow A^{2-} + H_2O$	Second deprotonation of analyte acid	$\xi_6 = n_{(AH^-)_{initial}} - n_{(AH^-)_{final}} + \xi_5$

By Hess' law, the net enthalpy due to these neutralization reactions, $\Delta H_{neutralization}$, is the sum of the products of the molar extent of each reaction, ξ_i, times the molar enthalpy of that reaction, ΔH_i:

$$\Delta H_{neutralization} = [\Delta H_1 \quad \cdots \quad \Delta H_6] \times \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_6 \end{bmatrix} \tag{61}$$

In this work enthalpies of neutralization for reactions 1-4 were supplied from the literature, [200, 213] and the enthalpies of reaction 5 (for FA, FFA, and FDCA) and reaction 6 (for FDCA only) were fitted. The supplied parameters are summarized in Table 16. The enthalpies of neutralization for citric acid were taken by Bernard and Burgot (1979) [200] (and are equal to values reported by Christensen, et al. (1967) [213] for the corresponding ionization reactions, plus the enthalpy of neutralization for water, also in accordance with Hess' law).

Table 16: Literature values for enthalpies of neutralization of water and citric acid.

Reaction No. from Table 15	$\boldsymbol{\Delta H_{neutralization}}$ **(J mol^{-1})**
1	-55740
2	-51566
3	-53313
4	-59087

5.3.5.6 Calculation of Dilution Enthalpies

Determining the enthalpic contribution due to the dilution and mixing of the solutes in the sample cell and ITC injection syringe is complicated by the very neutralization reactions discussed in the previous section. Tabulated data for heats of dilution are typically reported as relative apparent partial molal enthalpies of infinite dilution, L_ϕ, and include any heat effects that might be attributed to neutralization reactions such as those in Table 15. Plots of $L_\phi$vs. $\sqrt{m}$, where m is solution molality, are called "integral plots" in the nomenclature of Young and Vogel (1932) [195] and the vertical displacement of a chord along this curve beginning at $\sqrt{m_a}$ and ending at $\sqrt{m_b}$ is the apparent molal enthalpy $\Delta_{m_a}^{m_b}H_m$ of diluting a solution from initial molality m_a with relative apparent partial molal enthalpy $L_{\phi,a}$ to final molality m_b with relative apparent partial molal enthalpy $L_{\phi,b}$: [192, 195]

$$\Delta_{m_a}^{m_b}H_m = L_{\phi,b} - L_{\phi,a} = \Delta H_{dilution} + \Delta H_{neutralization} \quad (62)$$

Plots of experimental measurements of $\Delta_{m_a}^{m_b}H_m$ vs. $\sqrt{m}$ are commonly called "slope" or "chord" plots, given that they represent both the derivative and chords of the integral curve, L_ϕ. Unfortunately, these data do include enthalpies of neutralization, which have already been accounted for in this the overall model, and therefore the use of raw tabulated apparent dilution data could result in an unknown degree of double-counting for $\Delta H_{neutralization}$. Therefore, dilution data are needed from which the heat effects of neutralization have been removed. Surprisingly, we found only sparse examples of related

work in the literature, [196-198] and no tabulated adjusted dilution data from which estimated neutralization effects have been removed that were suitable for use in this study.

To avoid double-counting for $\Delta H_{neutralization}$, the enthalpies of neutralization included in the literature dilution data were estimated and subtracted to provide adjusted slope plots that were then used to calculate dilution enthalpies. Tabulated molalities [h] and L_ϕ values for each of the buffer components: citric acid, [192] NaOH, [111] NaCl, [111] and HCl [111] were used to initialize the model equations Equation (59)-(60) (in the absence of any other solutes or buffer components) to simulate the original (i.e., literature) dilution experiments. The reaction enthalpies were calculated according to Equation (61) and subtracted from $\Delta_{m_a}^{m_b} H_m$ according to Equation (62) to yield the dilution enthalpy $\Delta H_{dilution}$, which constitute an adjusted slope plot. These data were fitted to Equation (30), which is not of any theoretical basis, but empirically was found to capture the linear and exponential components of the data, where a,b,c, and d are fitted parameters:

$$\Delta H_{dilution} = a \times 10^{b \times \sqrt{m}} + c \times \sqrt{m} + d \tag{63}$$

The calculated values of $\Delta H_{dilution}$were numerically integrated over each dilution chord, according to the method established by Young and Vogel (1932) [195], and the fit of Equation (63) was analytically integrated, yielding two alternative integral plots. Vertical

[h] This study is limited to dilute electrolyte solutions, such that interconversion of molalities and molarities is a reasonable first-order approximation sufficient for demonstration of this method on our limited dataset. Therefore, the assumption of constant density was made and no further corrections were applied.

intercepts on plots of L_ϕ are arbitrary, and therefore integration constants were chosen to align the first point of the adjusted integral plots with the first point of the original dataset, facilitating visual comparison of the original and adjusted curves. The adjusted partial molal heats of infinite dilution, $L_{\phi,adjusted}$, were then used in the conventional manner to calculate the heats of dilution due to each individual solute in the sample cell between the states just prior to and immediately following titrant injection.

Dilution enthalpy of each analyte acid upon injection was accounted for by a slope parameter ω such that $L_{\phi,adjusted} = \omega\sqrt{C_{analyte}}$, which is an accepted limit for dilution enthalpies at sufficiently low solute concentration. [192, 195]

Binary interaction coefficients were also added to account for excess enthalpy, ΔH_{excess} due to the interaction of components that could not be fully captured by the $\Delta H_{neutralization}$ calculation:

$$\Delta H_{excess} = \sum_{i,j} \mathscr{b}_{ij} \left((C_i \times C_j)\big|_{final} - (C_i \times C_j)\big|_{initial} \right) \tag{64}$$

The five [i] symmetric binary interaction parameters, $\mathscr{b}$, for the interactions of the buffer components were fitted to the control dataset (i.e., the titrant was only water), and then held constant for subsequent analyses. The four additional parameters $\mathscr{b}$ for interactions of the

[i] i.e., (1) NaCl-NaOH; (2) NaCl-HCl; (3) NaCl-citric acid; (4) NaOH-citric acid; (5) HCl-citric acid. No parameter was included for the interaction of HCl-NaOH; equimolar amounts of HCl and NaOH were treated as NaCl such that an excess of only one or the other can exist in a particular sample at a time.

analyte with the buffer components, as well as the pK_a value(s) of the analyte were then also fitted for each other dataset separately. Fitting was performed with the nonlinear least squares method in MATLAB. Absolute average deviation (AAD) and absolute relative deviation (ARD) were calculated for the initial pH_c of each buffer, and for the integrated heats of injection for each titrant (water, FA, FFA, and FDCA) according to Equations (65) and (66).

$$AAD = \frac{1}{n}\sum_{i=1}^{n}|model\ value - measured\ value| \tag{65}$$

$$ARD = \frac{1}{n}\sum_{i=1}^{n}\left|1 - \frac{model\ value}{measured\ value}\right| \times 100\ \% \tag{66}$$

5.3.6 Results

5.3.6.1 Preparation of Titrand Solutions and Determination of Proton Concentration

31 Citric acid buffers were prepared that varied only in the total sodium and chloride content. As the charge balance in Equations (46), (58), and (59) demonstrates, excess of sodium is mathematically indistinguishable from titration with sodium hydroxide, and excess chloride is indistinguishable from titration with hydrochloric acid. The measured pH_c values of the 31 titrand buffers are presented in Figure 46, along with the total sodium and chloride concentrations of each buffer. Each buffer has a citric acid concentration of 20.0 mM and an ionic strength of 0.1. The buffers were prepared more densely between pH 2 and 4, and more sparsely at the periphery of the range, as shown.

Typical potentiometric pH determination yields an assessment of proton activity, $pH_a = -\log_{10}(a_{H^+})$, which differs from pH reported on a concentration basis, $pH_c = -\log_{10}[H^+]$. [214] Calibration of a standard electrode using the more involved method discussed by Gans and O'Sullivan (2000) [187] enables potentiometric measurements to be correlated to proton concentration directly. Electrode calibration yielded a Nernst slope of 58.63 ± 0.09 mV and a standard electrode potential of 373.0 ± 0.8 mV. pH_c values for each buffer were ultimately determined to within an uncertainty of 0.1.

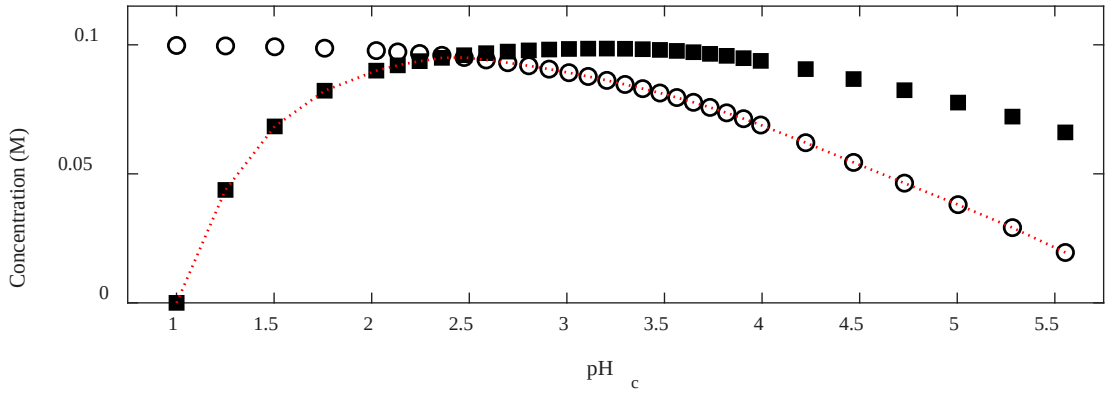


Figure 46: Composition and measured pH_c values of the titrands prepared in this study. Total concentrations of sodium ion (■) and chloride ion (○) in each of the 31 buffers, plotted against the resulting measured pH (concentration basis) measured as described below. The lesser number of the sodium and chloride concentrations for each pH value is the concentration of sodium chloride (·····); excess of either sodium or chloride is the concentration of sodium hydroxide or hydrochloric acid, respectively, that was added to each buffer

5.3.6.2 Spectrophotometric Determination of pK_a Values

The pKa values of FA, FFA, and FDCA were determined by spectrophotometry, which is a generally accepted method, [208, 215] to provide a basis of comparison for the pKa values data determined by calorimetry. The absorbance spectra of FA, FFA, and FDCA after buffer subtraction are shown in Figure 47, along with the deconvoluted absorbance spectra of each protonation state of each acid. FA has an isosbestic point near 247 nm, and FFA has two isosbestic points, near 242 and 282 nm. FDCA has three accessible protonation states, each with a unique absorbance spectrum, and therefore has no isosbestic point. Fitted concentration basis pK_a values and extrapolated thermodynamic values are reported in Table 17 on Page 178. The extrapolated pK_a^{th} for FA differs from the prior literature [112] value of 3.164 by 3 %. Experimental pK_a values for FFA or FDCA have not been reported previously.

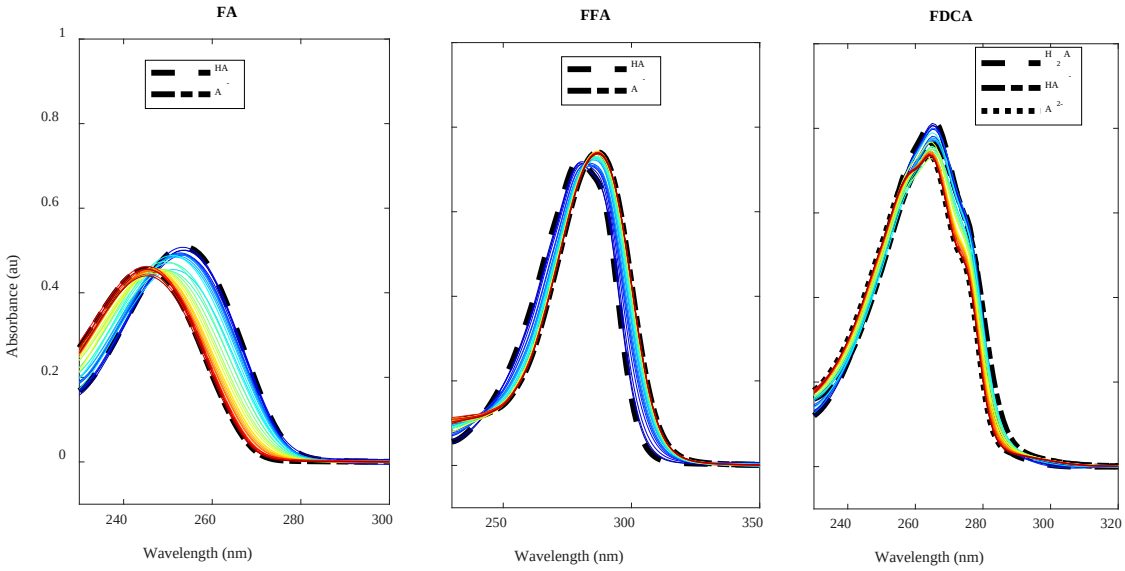


Figure 47: Halochromic shifts in the ultraviolet-visible absorbance spectra of FA, FFA, and FDCA. Each panel shows UV-Visible absorbance spectra of one of the furan-series acids in solution at 31 different pH values encompassing their pKas. More acidic samples are shown in blue and more basic samples in red. Dashed black curves indicate the deconvoluted absorbance spectra of each pure protonation state of each acid (i.e., HA is fully protonated mono acid and HA^- is singly deprotonated diacid). FA and FFA are monoprotic acids and each exhibit a single halochromic transition, whereas FDCA, a diacid, exhibits two transitions. FA has a single isosbestic point at 247 nm, and FFA has two isosbestic points, at 242 and 282 nm. FDCA has three accessible protonation states with different absorbance spectra, and therefore has no isosbestic point.

5.3.6.3 Removal of Reaction Enthalpies from Heats of Infinite Dilution

The experimental apparent molal enthalpies of infinite dilution, L_ϕ, from literature that were used in this study are plotted and referenced in Figure 48, along with the calculated relative molal enthalpic contribution due to neutralization reactions and the adjusted molal enthalpies of infinite dilution, $L_{\phi,adjusted}$. The enthalpic components due

to neutralization, which were calculated using the procedure described in section 5.3.5.6, were minor (i.e., flat and low in magnitude) relative to the dilution enthalpies for sodium hydroxide, sodium chloride, and hydrochloric acid, in the concentration range evaluated. These low magnitude values follow from the assumption that sodium hydroxide, sodium chloride, and hydrochloric acid are all strong electrolytes that are fully dissociated in aqueous solution at any concentration and therefore not influenced by dilution. This assumption is explicitly captured by the $[Na^+]$ and $[Cl^-]$ terms in the charge balance equation. In contrast, the calculated neutralization enthalpies for dilutions of citric acid, a weak acid with multiple accessible protonation states, were of the same order of magnitude as the reported apparent dilution enthalpies from the literature. Subtracting the calculated neutralization enthalpy from the literature data caused the shape of the $L_{\phi,adjusted}$ curve to deviate significantly from the original L_ϕ curve.

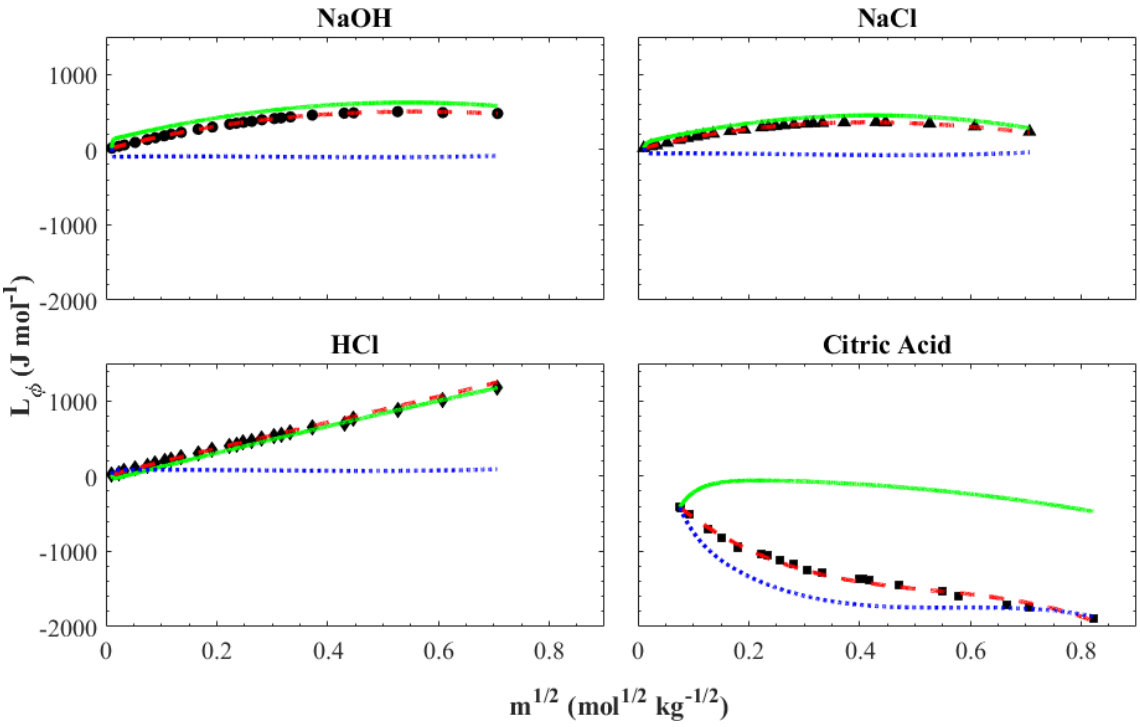


Figure 48: Apparent and adjusted molal enthalpies of infinite dilution, L_ϕ, for NaOH, NaCl, HCl, and citric acid. Apparent molal enthalpies of infinite dilution from the literature are displayed as filled markers. NaOH (●), NaCl (▲), and HCl (♦) data are from Wagman, et al. (1982) [111] and data from Dobrogowska, et al. (1990) [192] were used for citric acid (■). Fitted integral curves for each datasaet are plotted (– – ·). Calculated contributions of neutralization reactions to the apparent dilution enthalpy are shown for each solute (·····), and the final adjusted dilution enthalpies, $L_{\phi,adjusted}$, are also shown (——) on the same L_ϕ axis. Only the dilution enthalpies of the weak acid, citric acid, were drastically affected by this mathematical treatment.

5.3.6.4 Calorimetric Measurements

The integrated heat of injecting water into water was determined to be $-9 \times 10^{-7}\text{J} \pm 1 \times 10^{-7}\text{J}$ (mean of three measurements). This value was considered representative of the frictional heat of injection and mixing, and was subtracted from

subsequent experiments. The integrated heats of injection for each combination of titrant and titrand after subtracting this value are displayed in Figure 49. Unlike the ITC data collected by Tajc, et al. (2004) [205], Shoghi, et al. (2009) [206], and Zheng, et al. (2015) [207] which all display the expected characteristic sigmoidal curve of equilibrium-limited binding events, the data collected in this study exhibit unexpected behaviors. Most notably, the control experiment demonstrates that simply diluting citric acid buffers of the same concentration and ionic strength, but differing in pH value, could be either an endothermic or exothermic process. Furthermore, both the sign and magnitude of the net integrated enthalpy of the dilution vary continuously with the pH value of the solution in a non-monotonic manner that is difficult or impossible to intuit without modeling. All of the datasets in Figure 49 exhibit a peak in integrated heat of injection around pH 2.4. This pH value corresponds to the buffer to which the greatest total sodium chloride was added (as shown in Figure 46). As all buffers were prepared to achieve the same ionic strength, this is also the buffer in which (a) citric acid contributes the least to the overall ionic strength and (b) the sensitivity of ionic strength to changes in pH is minimized (as the slope of the red dashed curve in Figure 46 is shallowest at this pH value).

Integrated heats of injection for each of the other titrants exhibit perturbations from the trends of the control experiment. All three analytes introduce a trough in the integrated heat curves near pH_c 4, where the control experiment presents a shoulder, and coinciding with the mean of the second and third pK_a values of citric acid at 0.1 M ionic strength. In the case of FA and FDCA the trough appears to be "narrow," whereas it is wider and flatter in the case of FFA. FDCA, which has two labile protons, exhibits a deeper trough than

either FA or FFA. There are no obvious features that align on the pH_c axis with the pK_a values that were measured for each analyte acid by spectrophotometry, even if the water curve is subtracted from each of the other datasets.

In all four sets of experiments, the integrated heats of the three recorded injections (which were the second, third, and fourth titrant additions in the experiment, by convention) are tightly clustered, indicating that the dilution of the titrand and (excepting the control experiment) addition of the analyte acids influence each subsequent titration by less than the experimental error.

5.3.6.5 Fitting Integrated Heats of Dilution

The fitted curves for each injection of each titrant into each buffer are overlaid onto the experimental data in Figure 49, and the model does a very good job of capturing the graphical features of the experimental data that were discussed in the previous section. There are also trends in the fitted curves that are not apparent in the experimental data, such as that the peak enthalpy decreases slightly with each injection. Due to inherent experimental noise and variation, it is not possible to validate this trend using the experimental data. However, it stands to reason that heat effects dominated by the titrand (as the peak appears to be) would become less pronounced as the titrand becomes more dilute with subsequent injections.

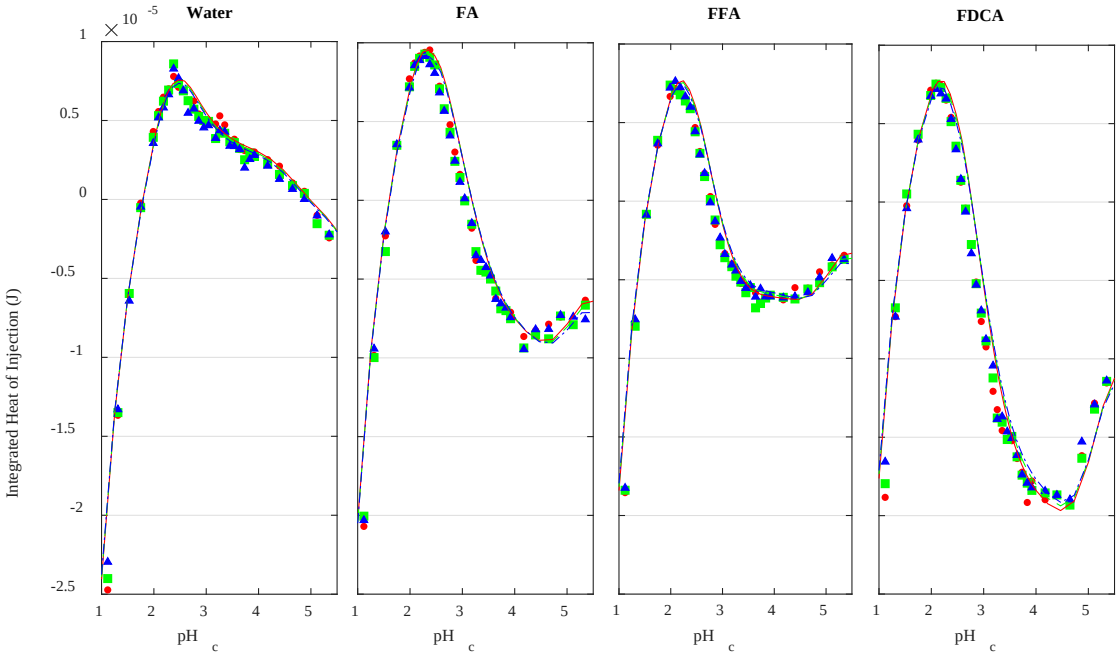


Figure 49: Integrated heats of injection and fitted model curves for the injection of water, FA, FFA, or FDCA into 31 citric acid buffers. Solid markers show the integrated heats of the second (•), third (■), and fourth (▲) injections, respectively, after subtracting the frictional contribution from all points uniformly. Each individual experiment is thus represented by a single cluster of one red, one blue, and one green marker on the same horizontal coordinate. Curve traces illustrate predictions from the fitted model corresponding to the second (———), third (– – – –), and fourth (–·–·–·–) injections. The absissca is the pH (on a concentration basis) of each buffer prior to the first injection, as determined by potentiometry after electrode calibration. The integrated heat of injection is the total heat that the calorimeter had to add to the sample cell in order to maintain the setpoint temperature of 25 °C. Therefore, negative values indicate that a titration was exothermic, while positive values indicate that a titration was endothermic.

Fitting the data also enables concealed behaviors to be deconvoluted and displayed. For example, the extents of each neutralization reaction, ξ, within the complex mixture are not values that can be probed with direct experiment. However, as part of the modeling process involves calculating these values, it is possible to infer what the data from such an

experiment might look like. Figure 50 shows, as a representative example, the calculated extents of each neutralization reaction that occurs for the second injection of FFA into each titrand, and the degree to which each reaction contributes to the total heat released or absorbed in a particular experiment. The extent of each neutralization can be seen to vary dramatically across the range of buffers, explaining, at least in part, the unintuitive shape of the experimental data.

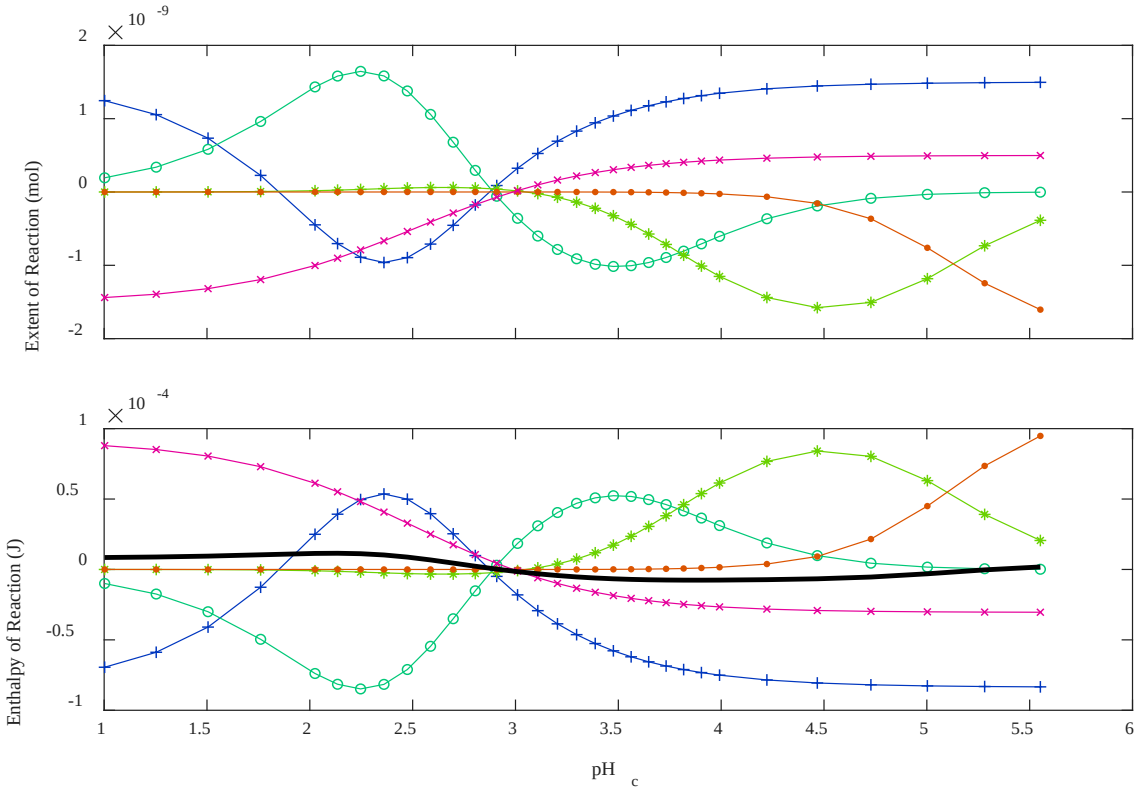


Figure 50: Extents and enthalpies of neutralization for the second injection of FFA titrant into each titrand buffer. The top and bottom plots show the molar extents, ξ, and net enthalpies, ΔH, for each neutralization reaction: $H^+ + OH^- \rightarrow H_2O$; $H_3B + OH^- \rightarrow H_2B^- + H_2O$; $H_2B^- + OH^- \rightarrow HB^{2-} + H_2O$; $H_2A + OH^- \rightarrow HA^- + H_2O$; $HA^- + OH^- \rightarrow A^{2-} + H_2O$. The total enthalpy of neutralization, $\Delta H_{neutralization}$, is also shown in the bottom plot (———). Enthalpies of each neutralization are the products of the extents and molal enthalpies of each neutralization reaction, so the data in the bottom plot are transformations of the corresponding series in the top plot. All data are plotted against the initial pH_c of the titrand buffers prior to titrand adddition.

The fitted pK_a^{th} values and their uncertainties are reported in Table 17. The pK_a^{th} values for FA and FFA are within reasonable agreement with those determined by the spectrophotometric method, while those for FDCA differ by 32% and 31% for the first and

second protonation steps, respectively. Generally speaking, the pK_a values determined spectrophotometric method are probably more reliable than those determined using this ITC approach for several reasons. Spectrophotometry provides a fairly direct observation of the state of the analyte acids and only minimal deconvolution is required. Furthermore, the influence of the background citric acid on measurements is reduced by its low absorbance at the peak absorbance wavelengths of the analyte acids. Conversely, it can be seen from the relative magnitude of the total enthalpy of neutralization, $\Delta H_{neutralization}$, curve in the bottom panel of Figure 50 that the net measurable heat is very small (in magnitude) relative to the heats of ionization of each individual component. The charge balance constraint effectively forces the autoionization of water to counteract most of the measurable heat signal, reducing the sensitivity of the method. Finally, a great deal of deconvolution is required to interpret the calorimetric data Each step in this interpretation relies on additional empirical data, such as the relative apparent molal heats of dilution, L_ϕ, and the ionic strength correction (Equation (60)) which uses empirical parameters that were determined by regression of data tabulated in yet another source (i.e., the 1975 version of the Smith, et al. (2004) [211] database) which contains a broad range of empirical data for other acid complexes, although the furoic acids used in this study are not among them.

Additional error may be engendered specifically in the case of FDCA as its second pK_a^{th} value (3.44 by spectrophotometry) is closely bounded by the first and second pK_a^{th} values of the background buffer, citric acid (2.859 and 3.880, respectively). There may be other configurations of this experiment which could avoid this issue, such as titrating FDCA into alternative buffer or salt solutions.

Table 17: pK_a Values Determined for FA, FFA, and FDCA by UV-Visible Absorbance and Isothermal Titration Calorimetry. Values are fitted parameters and are listed with their calculated uncertainty.

	Spectrophotometric Method				ITC Method	
	$\boldsymbol{pK_{a,1}^{c}}$	$\boldsymbol{pK_{a,2}^{c}}$	$\boldsymbol{pK_{a,1}^{th}}$	$\boldsymbol{pK_{a,2}^{th}}$	$\boldsymbol{pK_{a,1}^{th}}$	$\boldsymbol{pK_{a,2}^{th}}$
FA	2.860 $\pm$0.005	NA	3.061 $\pm$0.005	NA	2.99 $\pm$0.02	NA
FFA	2.015 $\pm$0.003	NA	2.216 $\pm$0.003	NA	2.31 $\pm$0.03	NA
FDCA	1.86 $\pm$0.04	3.03 $\pm$0.01	2.06 $\pm$0.04	3.44 $\pm$0.01	2.72 $\pm$0.02	4.5 $\pm$0.3

pK_a^c: pK_a value measured and reported on a concentration basis at an ionic strength of 0.1 M.
pK_a^{th}: pK_a value extrapolated to a thermodynamic basis.
NA: Not Applicable.

The binary interaction coefficients, $\mathscr{b}$, are reported in Table 18; those parameters between each of the titrand components (i.e., sodium chloride, sodium hydroxide, hydrochloric acid, and citric acid) were determined in the first fitting operation which involved using only the control dataset. The remaining fitted parameters are specific to each analyte, and were fitted in subsequent rounds of analysis while holding the binary interaction coefficients of the titrand components to be fixed parameters.

Table 18: Fitted binary interaction coefficients. Each cell contains the binary interaction coefficient, $\mathscr{b}_{i,j}$ ($\mathbf{J\,L^2\,mol^{-2}}$), and uncertainty to capture the specific enthalpic effects between the solutes listed in the corresponding row and column heading.

		Solute 1			
		NaCl	**NaOH**	**HCl**	**Citric Acid**
Solute 2	NaCl	NA	NA	NA	NA
	NaOH	0.004 ± 0.002	NA	NA	NA
	HCl	-0.002 ± 0.001	NA	NA	NA
	Citric Acid	0.0046 ± 0.0007	0.003 ± 0.004	0.002 ± 0.002	NA
	FA	2.8 ± 0.5	4 ± 1	2.5 ± 0.5	-14 ± 3
	FFA	3.2 ± 0.4	4.6 ± 0.8	3.2 ± 0.4	-16 ± 2
	FDCA	5 ± 1	8 ± 2	5 ± 1	-25 ± 6
NA: Not Applicable.					

The slope parameters, ω, and neutralization enthalpies $\Delta H_{neutralization}$ are presented in Table 19. The $\mathscr{b}$ parameters represent the excess enthalpies of mixing of the components that is due to their binary interactions in solution, rather than to dilution of the pure components or their influence on the extents of neutralization reactions. Based on the formulation of Equation (64), positive values of $\mathscr{b}_{i,j}$ imply that heat is absorbed when the product of the concentrations of species i and j decreases during an injection, while negative values imply that heat is released. In general, the interaction coefficients of the titrand components were lower in magnitude and less tightly fit (gauged by uncertainty as a percentage of the fitted value) than the interaction coefficients between the analytes and the titrand components. All analytes produced negative $\mathscr{b}$ values for their interactions with citric acid and positive $\mathscr{b}$ values for their interactions with the other titrand components. All fitted dilution slope parameters, ω, were positive, indicating that $L_{\phi,adjusted}$ for each

analyte is initially upward sloping, which is consistent with that for citric acid, the other weak acid in this study.

The fitted neutralization enthalpies for FA and FFA are negative and greater in magnitude than that for water, indicating that the ionization of FA and FFA are both exothermic – similar to, but greater in magnitude than the last ionization step of citric acid. The fitted neutralization enthalpies for FDCA are of questionable veracity, with the first suggesting an ionization enthalpy of more than double the magnitude of that of FA, and the second indicating an endothermic ionization step. It is unsurprising that these values are questionable, given that the fitted pK_a^{th} values for FDCA differed so greatly from those identified using the spectrophotometric method.

Table 19: Fitted dilution and neutralization enthalpy parameters and their uncertainties.

	Dilution Slope Parameter $\boldsymbol{\omega}$ $\left(\mathbf{J\ mol^{-3/2}\ L^{1/2}}\right)$	**Neutralization Enthalpy,** $\boldsymbol{\Delta H_{neutralization}}$ $\left(\mathbf{J\ mol^{-1}}\right)$	
		First Deprotonation Step	**Second Deprotonation Step**
FA	18000 ± 3000	-62000 ± 300	NA
FFA	26000 ± 2000	-61100 ± 200	NA
FDCA	102000 ± 4000	-68600 ± 500	-54000 ± 1000
NA: Not Applicable.			

The average absolute deviation and average relative deviation for each fit, as calculated according to Equations (65) and (66) are presented in Table 20. Typically, these values might be used as catch-all representations of how well the model represents each

dataset. However, such interpretations can be misleading, as illustrated by the ARD for FDCA which is low in comparison to those for water, FA, and FFA, even though the fitted parameters for FDCA arguably do a poorer job of capturing the reality of the physical mechanisms to which they correspond.

Table 20: Average absolute deviation (AAD) and average relative deviation (ARD) for the fitted calorimetry model. AAD and ARD were calculated according to Equations (65) and (66), respectively.

Fitted Value	$\boldsymbol{AAD}$	$\boldsymbol{ARD}$ (%)
pH_c	0.102	3.52
Integrated Heat; Water	2.82×10^{-7} J	26.4
Integrated Heat; FA	3.30×10^{-7} J	23.3
Integrated Heat; FFA	2.53×10^{-7} J	11.7
Integrated Heat; FDCA	3.85×10^{-7} J	10.4

5.3.7 Discussion

The specific data collected, fitted, and discussed within this work are perhaps of secondary importance to Equation (59) which encompasses the linearization of – and application of an eigenvalue problem statement to – calculations of pH and ionic equilibria. This approach has not been described previously, as recently published tomes dedicated to related ionic systems treat similar calculations as more difficult nonlinear problems. [188, 212]

It is unlikely that this method of applying ITC will soon overtake existing methods for pK_a determination, given that the uncertainties of the fitted pK_a values for FA and FFA were about tenfold those determined using the spectrophotometric method, and especially given the questionable results for FDCA. However, with more concentrated solutions and

more method and model development, it may take its place as another tool in the analytical chemist's toolbox. The method presented here for removing the contribution of neutralization reactions from existing literature data for partial molal enthalpies of infinite dilution is also a useful application of the math presented here. It is unclear whether such calculations are desired by the thermodynamics community, given that the previous approaches to calculating neutralization enthalpies in infinite dilution data are few in number and do not seem to have gained widespread usage. [196-198] $L_{\phi,adjusted}$ curves might not be missed if their absence can be masked by adjustable parameters. In our investigation, we found that the binary interaction coefficients, $\mathscr{b}$, could compensate for the use of uncorrected dilution data, taking on different values but yielding essentially the same cumulative AAD and ARD over all four fitted datasets. Fundamentally, however, aqueous weak acid dilution data are either inclusive of neutralization or they are not. This method or similar techniques may be retroactively applied to other dilution data thanks to the Chord method of Young and Vogel (1932) [195].

Overall, the data presented here help to demonstrate how this refined approach to handling ionic equilibria can be applied to a variety of problems, from designing buffers with desired properties as in Figure 46, to breaking down thermodynamic data as in Figure 48 and Figure 50. Furthermore, since the kinetics of ionization and neutralization reactions occur much more rapidly than other reactions in aqueous solution, [216-218] this method for calculating the equilibrium pH of a system could be incorporated into other models to provide the "instantaneous" pH value of even dynamically reacting chemical systems. The model presented here incorporates a finite-difference discretization of the CSTR mass-

balance (Equation (37)), and it is reasonable to imagine reverting this calculation to a time-step rather than a volume-step and performing many numerical integrations, (rather than just the four injections calculated here) to capture the transient and steady state behaviors of continuously operated reactors with pH-dependent reactions in industrial chemical processes. An example of how this approach can be applied to batch or plug flow reactors with series reactions is provided in CHAPTER 6, section 6.2.3.3, which discusses possible future directions for reaction engineering of the valorization of furfural to FDCA.

CHAPTER 6. CONCLUSIONS

6.1 Broader Perspective

6.1.1 Looking for Catalysts in Nature

In a recent Origins of Life seminar at Georgia Tech, [219] Dr. Robert Hazen posed an intriguing Fermi question: *How many mineral-catalyzed Miller-Urey type [220] origins of life chemistry experiments have been performed by the planet Earth since it cooled?* Hazen proposed that a reasonable estimate might be the product of the planet's mineral surface area and the time since the formation of the moon, divided by the product of the average time and surface area required for a chemical reaction to take place. By Dr. Hazen's estimation, shown in Equation (67), the planet has performed on the order of 10^{53} origins of life chemistry experiments catalyzed by mineral surfaces. [221]

$$\frac{\begin{pmatrix} Mineral \\ surface\ area \\ of\ Earth \\ \approx 10^{28}\text{cm}^2 \end{pmatrix} \times \begin{pmatrix} Time\ between \\ cooling\ of\ Earth \\ and\ earliest\ fossils \\ \approx 2 \times 10^{16}\ \text{s} \end{pmatrix}}{\begin{pmatrix} Surface\ area \\ required\ for \\ typical\ reaction \\ \approx 10^{-12}\ \text{cm}^2 \end{pmatrix} \times \begin{pmatrix} Duration \\ of\ a\ typical \\ reaction \\ \approx 10^{3}\ \frac{\text{s}}{\text{experiment}} \end{pmatrix}} = \begin{pmatrix} Surface \\ catalyzed \\ prebiotic \\ chemistry \\ experiments \\ \approx 2 \times 10^{53} \end{pmatrix} \quad (67)$$

The aspect most important to this discussion is not what this number is precisely, but only that the number incomprehensibly large. The key point is that the planet itself is a prolific experimenter, and is much more likely to stumble upon statistically rare (i.e.,

interesting) results than any graduate researcher, or even *every* graduate researcher. In fact, if every human who has ever been born ($\approx 1.08 \times 10^{11}$, according to Haub (2011) [222]) spent five years in graduate school performing a catalysis experiment every day, they would have collectively performed about 2×10^{14} experiments; a paltry duodecillionth ($\approx 10^{-39}$) of the research and development output of the earth itself before life even began. And that is a particularly generous estimate which assumes that research funding is readily available to all 10^{11} graduate students!

A similar question, and one of more direct relevance to this thesis, could be posed in relation to research and development of enzyme catalysts. Enzymes are generally more selective and specific than other catalysts, and are of interest for their application to substrates that are biologically sourced (such as plant biomass), or to products that must interact with biology (such as pharmaceuticals) due to the conserved themes of shape and chemical functionality that define the bioorganic chemistry space. Unfortunately, our understanding of proteins and enzyme catalysis is not yet to the point where we can reliably engineer desired catalytic activities *de novo*. And as estimated above, there are hard limits on the number of brute-force experiments that researchers can perform to develop biocatalysts with desired activities.

Bacteria, like the Earth, are productive experimentalists. An estimated 2×10^{30} new prokaryotic cells are formed each year. [223] Errors in DNA synthesis arise at different rates in different bacterial strains and species, but generally occur at about 0.0033 per replication event, with roughly two thirds of those localized to protein-encoding regions. [224] If 16 % of protein-coding regions correspond to enzymes, [225] then bacteria

collectively develop and test about 7×10^{26} new enzymes each year, almost four trillion times more than our generous estimate for the cumulative research power of all of humanity for all of time.

This mental exercise leads to a simple but powerful idea: bacteria have us at a staggering disadvantage. Strategic planning would suggest that – at least until we better understand how to predict enzyme function – the most efficient paradigm for biocatalyst development and bioprocess engineering is one in which we work in deference to nature. Staying true to this philosophy, this project sought to find enzymes with demonstrated activity on the target molecules (furfural, FA, FFA, and FDCA) within existing metabolic pathways.

6.1.2 Characterization Before Application

The research paradigm underlying this study allows important applications to direct the allocation of resources and may help to reduce expenditure on uninteresting or barren research topics. Unfortunately, there is a caveat to this approach – a potential trap for unwary engineers. Most systems that are engineered *de novo* and built from the ground up are well-understood; every truss in a bridge or cam in a piece of manufacturing equipment has its purpose. In application-driven projects employing biocatalysis, however, the need to pull from biology may lead to situations in which new discoveries from the natural world are thrust into use before they are adequately understood. When this occurs, the apparent initial benefits of pulling from nature may be counterbalanced by the downstream costs of having bypassed fundamental scientific characterization. As engineering researchers find

they are building on shaky ground, they may need to backtrack and reallocate resources to generate foundational knowledge.

Ultimately, we were unsuccessful in developing a wholly enzymatic route from furfural to FDCA. We did not find this to be due to any fundamentally insurmountable challenges (e.g., impossible chemistry or violations of thermodynamics), but only because the basic building blocks were not yet at a sufficient stage of development to allow them to be linked together into a process. The enzymes we selected as catalysts are not yet understood. In fact, one series of studies published concomitant to this this project [91-93, 95-96] was enough to entirely redefine what was known about the HmfF/HmfG enzyme families. Unfortunately, due to our inability to successfully stabilize HmfF and HmfG, we could not really apply this new information, and most of the effort in biocatalyst development went towards testing of different purification strategies. Purification of stable protein is not a fundamental requirement for characterization, but a practical one.

At the start of this work, furfural, FA, FFA, and FDCA were also not sufficiently understood to inform process development decisions. There was not enough physical property data to enable reliable thermodynamic predictions or estimates of limiting conversion. Physical property studies were conducted in this study so that these very calculations might be carried out at a later date. Even the tools available for predicting general interactions of organic acids – such as how pH of a solution might change over the course of a reaction – were found to be limiting. The details of charge balances and activity coefficients might have been picked up and correctly applied by thermodynamicists and

synthetic chemists, but there has been only limited return of information to the biology, biochemistry, and protein communities for whom the pH scale was initially conceived. [226]

6.1.3 Driving Interdisciplinary Collaborations for Research

In general, these types of issues can be addressed by actively working to foster greater communication between the scientific disciplines, which otherwise grow increasingly specialized. [227] Platforms from which both applied and fundamental researchers discuss what they need from – or can offer to – their counterparts are necessary to bridge this gap, and have been crucial to the development of this project. Conferences with attendance from scientists and engineers with disparate backgrounds are great ways to kickstart collaborations and the sharing of information and materials. For example, representation from our research group at the 19th International Symposium on Flavins and Flavoproteins (July 2-6, 2017, Groningen, The Netherlands) led to a collaboration in which the Leys Research Group (University of Manchester, Manchester, UK) provided us with their expression construct for UbiX. The Atlanta Flavin Group, which primarily consists of enzymologists and biochemists welcomed us into monthly meetings, which led to a lot of shared information on the tools and techniques involved in fundamental enzyme characterization. Tools, guides, and review articles focused on making one discipline accessible to another [137] are critical to broadcasting opportunities for collaboration, such as the use of PC-SAFT in this work.

6.2 Recommended Topics for Further Study

This project has provided context and groundwork for a plethora of cross-disciplinary follow-up studies with both fundamental and applied significance. The enzymes discussed in this study could lead to the first industrial use of cofactors and chemistry that were only described less than three years prior to this writing. [91-92] There are opportunities remaining in physical property characterization and modelling for aqueous systems of furfural, FA, FFA, and FDCA, and there are opportunities to use those tools to inform process development and reaction engineering. To that end, the following sections provide a number of suggestions for additional cross-disciplinary investigations that could be performed to continue the development of this project in the areas of enzymology and protein engineering, thermodynamics and modelling, and reaction engineering and process development.

6.2.1 Future Work in Enzymology and Biocatalyst Development

6.2.1.1 Characterization of HmfF and HmfG

In relation to enzymes as "new" as HmfF and HmfG, it is difficult to imagine a limit to the possible avenues that may be opened by future research. Studies should be directed in the short-term to gain more insight into the reversible C-CO_2 bond cleavage by HmfF and similar enzymes, to probe the mechanism of flavin transfer between HmfG and HmfF, and to identify structural features of each protein that are important to their respective mechanisms for comparison to UbiX and Fdc1.

Without crystallographic structural information (as is currently the case for HmfF and HmfG), bioinformatics and consensus-statistics tools can be used for rational design. As stable purified enzyme has been difficult to obtain, protein stability could be a fruitful first design target, as a stepping stone to kinetic and mechanistic studies. Stabilizing mutations may make HmfF and HmfG much more amenable to purification and facilitate further study. The frequency with which a particular amino acid is found at a particular position across the population of homologues can be used to predict mutations that may increase the energetic penalty of unfolding, $-\Delta G_{folding}$, according to the Boltzmann distribution of the canonical ensemble of possible mutations at a given position. [228-229] The general impact of stabilizing mutations can be evaluated by observing changes in the melting and aggregation onset temperatures by DSF, and by measuring how well enzyme variants retain their activity after incubation in a variety of storage conditions. Specific impacts of stabilizing mutations on enzyme conformation can be evaluated by temperature-scanning circular dichroism (CD) spectroscopy. Finally, net changes in unfolding energy can be evaluated by differential scanning calorimetry (DSC).

As an alternative to stabilizing HmfG, additional studies are already being conducted to couple the prenyltransferase activity of UbiX with HmfF. If pure and active holo-HmfF can be prepared with prFMN produced by UbiX, as has been demonstrated with Fdc1,[92] it will enable mechanistic studies of HmfF without necessarily requiring detailed study of HmfG.

The conjugated structure of the flavin cofactor enables structural and mechanistic information to be gathered by techniques that probe the resonance states of the cofactor,

such as spectroscopy, fluorescence, and electron paramagnetic resonance. Spectroscopy with stopped-flow instruments can provide detailed rapid kinetic information that can identify short-lived structural intermediates and provide insight into the catalytic properties of the flavin prenyltransferase and flavin-dependent decarboxylase enzyme families.

Pull-down assays, such as by purification of affinity-tagged HmfF that has been coexpressed with untagged flavin prenyltransferase (either HmfG or UbiX), or vice versa can be used to probe physical interactions between the two proteins. Additional research tools that can be used for this purpose are size exclusion chromatography with multi-angle light scattering (SEC-MALS) and analytical ultracentrifugation (AUC) which can both also be used to evaluate the multimeric states of the individual enzymes.

As of this writing, X-ray crystallography is the established method for determining protein structure. Although there is growing debate as to how representative crystal structures are of active solvated proteins, crystallography is the only method with which to obtain detailed information on the positions of side-chains and domains within proteins. These structures are prerequisite to most targeted protein engineering tasks (i.e., anything beyond: the addition of simple tags, mutations based on sequence statistics, or directed evolution). Kinetic crystallography, single-crystal spectroscopy, and time-resolved single-particle cryoelectron microscopy can shed light on reaction mechanisms by tracking reactions and showing the relative alignments of active site residues, cofactors, and reaction substrates or products.[230-234] Further, it may be useful to investigate related substrates and active-site-directed mutations with the specific goal of poisoning catalysis by irreversible or kinetically-trapped formation of adducts. Crystallography of these states

can indicate specifically how – and in what order – interactions between active-site residues, cofactor, and substrate occur. [91] This type of information may help to elucidate chemical mechanisms responsible for C-CO_2 bond formation on furfural, as well as the structural difference within the UbiD family that are responsible for substrate specificity.

Future reaction engineering studies may be found to suggest better process performance in reactors with high partial pressure CO_2 and/or oxygen. If such a reactor design is to be evaluated, it will be necessary to determine how the carboxylase and oxidase (and possibly catalase and/or superoxide dismutase) enzymes perform under pressure. The effects of high hydrostatic pressure on enzymes have been explored previously. Reactions carried out under high hydrostatic pressure in deuterated solvent or with isotope-labelled substrates may be used to infer how each kinetic mechanism is affected. [235] The addition of polyols such as glycerol to the reaction medium may be found to provide a stabilizing effect against pressure deactivation, and elevated pressure may also confer some additional degree of thermal stability to the enzyme catalysts. [236]

6.2.1.2 Suggested Targets for Rational Design

With enough structural and mechanistic information, many other alluring protein engineering and rational design targets become viable research projects, and the initial goal of this thesis – upgrading furfural to FDCA – can be revisited. An HmfF-HMFO fusion protein with a flexible linker or DNA scaffold could be developed to reduce diffusion limitations by decreasing the distance FFA would have to travel before being oxidized to accelerate the process of oxidative product removal from the equilibrium-limited

carboxylation reaction. [237-238] Site-directed mutagenesis in the HmfF active site may be used to enhance the carboxylase activity on furfural, the industrial target substrate considered within this thesis. Catalytic efficiency of mutants can be assessed either by the real-time colorimetric pH-based assays described in CHAPTER 5, section 5.2, A High-Throughput pH-Based Colorimetric Assay: Application Focus on Alpha/Beta Hydrolases, or with endpoint assays, such as by chromogenic aldehyde functionalization with 4-hydrazinobenzoic acid (4-HZBA) or 4-hydrazino-7-nitro-2,1,3-benzoxadiazole (NBD-H). [239-240] Such functionalization can also be used to improve upon the HPLC methods that were used in this work. [241-244] Oxidation progress, and the effect of mutations or changes in reactor conditions on the activity of HMFO may be tracked by the methods listed above. Alternatively, HMFO activity may be evaluated indirectly with an amplex red and horseradish peroxidase assay that tracks the formation of hydrogen peroxide. [245]

The series reaction scheme presented in Figure 8 requires that both HmfF and HMFO be present in the reaction vessel at the same time – or perhaps even tethered as mentioned previously. This means that both enzymes would be in regular contact with furfural, FFA, and FDCA. As presented in Figure 15, the HMFO variant from Dijkman, et al. (2015) [71] was selected for the oxidation step of this route specifically because it is active only on FFA. HMFO has been reported to exhibit substrate inhibition, [71] and may also prove to be inhibited by furfural. Ligand-docking simulations such as those that yielded the original HMFO variant could be performed to identify potential mutations that might reduce substrate inhibition and possibly also inhibition by furfural, if it proves to be a problem.

HmfF is believed to be a native FDCA decarboxylase, and will likely be just as active (or more so) on FDCA as on furfural. If it is found that the reaction conditions that maximize carboxylation of furfural are insufficient to entirely prevent the decarboxylation of FDCA, then valuable product will be degraded into FA, as shown in Figure 51. This suggest that it may be worthwhile to seek HmfF variants that carboxylate furfural but are inactive on FDCA. FDCA is larger than either furfural or FA, so it may be possible to block its decarboxylation by constricting the active site of HmfF.

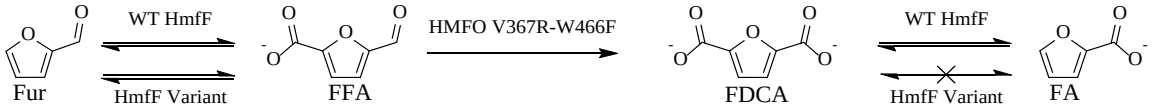


Figure 51: Reaction engineering suggests targets for future biocatalyst development. Wildtype (WT) HmfF may lead to product loss in a bioreactor designed to upgrade furfural to FDCA. The native activity of HmfF is decarboxylation of FDCA to FA. It may be possible to modify the HmfF active site to block FDCA decarboxylation, while still permitting carboxylation of furfural.

In addition to potential decarboxylative losses, the one-pot synthesis may be at risk of condensation losses. Furfural is thought to undergo oxygen radical mediated polymerization in aerobic environments, and the oxidation of FFA requires the presence of oxygen in solution. As shown in figure Figure 52, molecular oxygen can abstract a proton at the C5 position of furfural, leading the formation of a furan polymer. [246] It is possible that the formation of hydrogen peroxide by HMFO may even accelerate this process. It should be relatively straightforward to determine whether condensation polymerization is occurring, as the conjugated structure of the polymer results in a very high molar extinction

coefficient. Assessing the degree to which furfural is lost may be more difficult, as the optical absorbance of the polymer is likely a function of its length. Stopped-flow studies and mass spectrometry may be required to quantify furfural loss. Nevertheless, kinetic studies should be performed to compare the rate at which HmfF carboxylates furfural to the rate at which furfural polymerizes. It may be necessary to develop enzyme mutants specifically to avoid this condensation loss. Mutations to enhance the catalytic activity of HmfF on furfural and accelerate carboxylation may allow the desired reaction pathway to be more kinetically competitive with the undesired pathway.

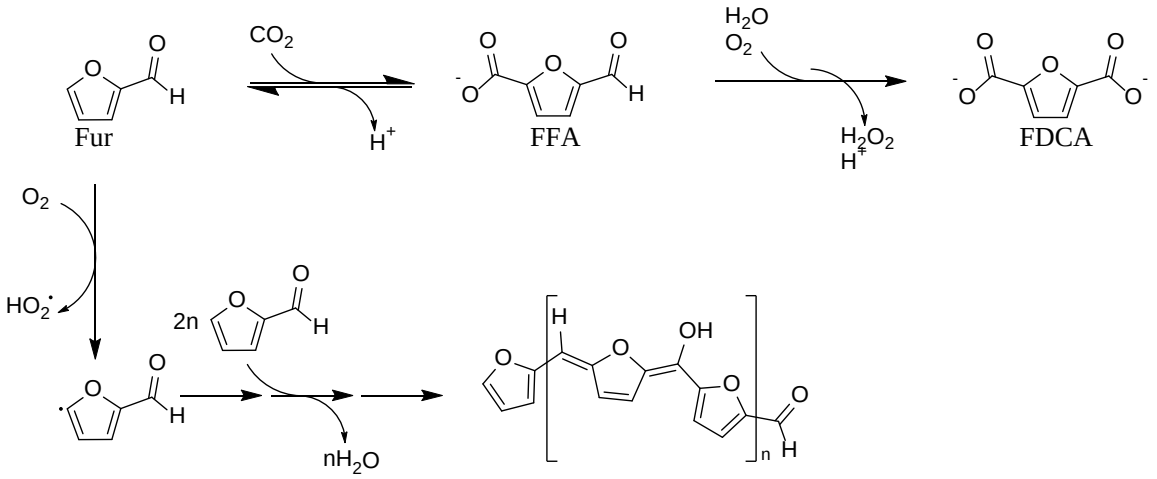


Figure 52: Fast carboxylation may be required to reduce condensation losses. Molecular oxygen can activate furfural leading to the formation of a polymeric byproduct and, eventually, a gel. [246]

If hydrogen peroxide is found to accelerate furfural condensation, it may be reasonable to introduce superoxide dismutase and or catalase enzymes into the reaction scheme to block radical formation. Oxidase-catalyse fusion constructs could also be

potential targets for protein engineering. It may even be possible to combine enzyme immobilization with substrate channeling and confinement by encapsulating HmfF, HMFO, and catalase into self-assembled protein nanoparticles. [247]

6.2.1.3 Opportunities for Development of Whole-Cell Biocatalysts

If holo-form HmfF cannot be stably purified, or if UbiX is found not to be a suitable replacement for HmfG, then free-enzyme catalysis may not be feasible for the carboxylation of furfural and metabolic engineering should be used to prepare a pathway comprising HmfF, HmfG or UbiX, and HMFO. Experimentation with different promoter strengths and media supplementation may be necessary to achieve high expression and sufficient flux. If engineered bacterial strains are found to survive on minimal media with only furfural, FA, FFA, or FDCA as the carbon source, then it may be necessary to knock-out any possible endogenous expression of homologues to the *C. basilensis* enzymes HmfA-D (Figure 11). This should help to prevent the whole-cell biocatalyst from simply eating the feedstock. Media supplementation of 2-oxoglutarate may become necessary if these gene deletions are found to have broader influence on overall metabolism.

6.2.2 Future Work in Thermodynamic Characterization and Modelling

The thermodynamic characterization conducted in CHAPTER 4 is only a subset of the work necessary to build a strong foundation for reaction engineering. The inorganic carbon source required for carboxylation is most likely bicarbonate, which is the dominant state of inorganic carbon around pH 8.5 (Figure 53), although it may also be influenced by solution temperature and CO_2 partial pressure. [190, 248]

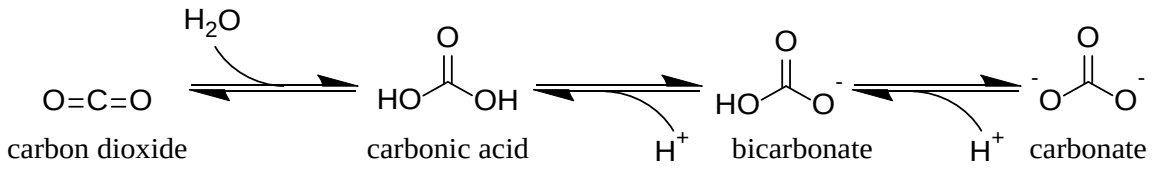


Figure 53: Aqueous carbon dioxide exists in equilibrium with carbonic acid, bicarbonate, and carbonate.

The modelling performed in CHAPTER 4 is limited to low pH values, due to the consideration only of saturating acids. However, the charge balance calculations discussed in CHAPTER 5 suggest that the addition of strong acids or bases to saturated solutions will cause the equilibrium composition of precipitated acid deposits to shift, and at the high pH values where bicarbonate is present, it is the solubilities of the sodium salts, not the protonated acids that may limit the phase space available to reacting systems (Figure 54).

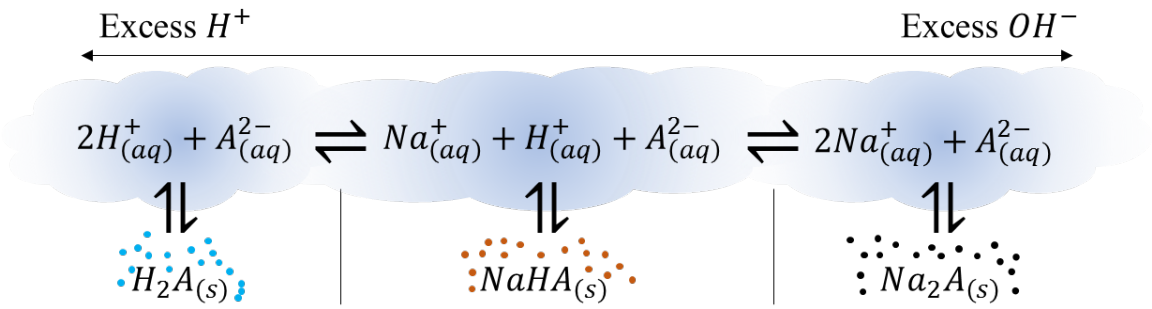


Figure 54: Proposed scheme for exchange of acid between aqueous and solid states. as sodium hydroxide is added to a solution saturated with a weak diprotic acid.

Additional characterization should be performed to investigate the occurrence of hydrates, coprecipitation of acids, or inclusion of furfural into the solid phase. Tools and techniques that might be leveraged for such a study include analytical titration, HPLC, X-ray powder diffraction (XRD), and Fourier transform infrared spectroscopy (FT-IR). [249]

The general modelling approaches of CHAPTER 4 and CHAPTER 5 could be combined to capture the equilibrium dynamics of pH, solubility, and thermodynamic activity over a broad pH range covering all acid protonation states, such as in the scheme proposed for future investigations in Figure 55. An extension of the PC-SAFT model that was used in this work, the *electrolyte* Perturbed-Chain SAFT (ePC-SAFT) might be of utility in this regard, as it incorporates ion-pairing corrections that are essential to capture some of the nonidealities of more complex electrolyte solutions.[250-252] Some similar features to those in the proposed solubility scheme have been described with PC-SAFT and ePC-SAFT previously, with pH as a model input. [155, 253] The techniques of CHAPTER 4 could ostensibly extend this type of modelling with the ability to predict the final solution pH.

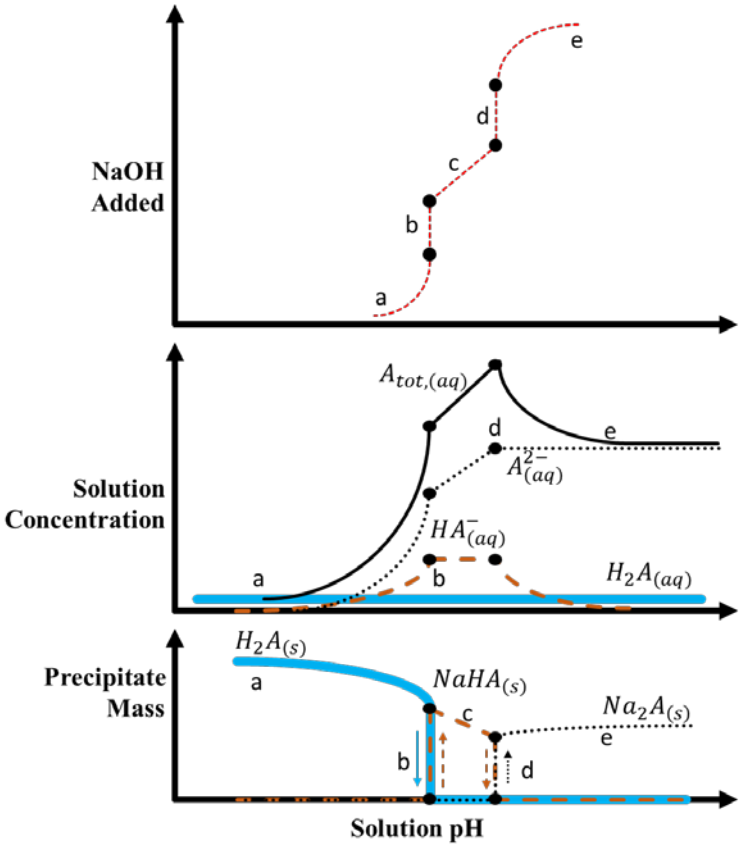


Figure 55: Proposed equilibrium dynamics between pH, acid solubility, and strong base for investigation in future modeling work. (a) Sodium hydroxide (-----) is added to a solution saturated with a diprotic weak acid (▬) causing the pH to increase. This in turn causes the fraction of acid in the monoprotic state (▬ ▬) to increase. (b) The monoprotic acid reaches its solubility limit and precipitates as monosodium salt. Microreversibility brings diprotic acid from the solid phase into solution, where it deprotonates and re-crystallizes as monosodium salt. Every additional sodium ion introduced to the solution brings a monoprotic acid molecule into the solid phase, and pH is fixed until no diprotic acid remains in the solid phase. (c) pH increases further, causing the monoprotic acid in solution transition into the doubly-deprotonated state (.....). (d) The fully deprotonated acid accumulates up to its solubility limit and begins to precipitate as a disodium salt. Again, pH is fixed as the addition of more sodium causes monosodium salt to resuspend, and then re-precipitate as disodium salt. (e) Once the reservoir of solid monosodium salt is depleted, the pH begins to rise again, causing the remaining monoprotic acid in solution to fully deprotonate. This causes the total acid in solution (▬) to pass through a maximum and dip back down to the solubility limit of the disodium salt.

Long-term modelling goals should be the ability to predict the equilibrium pH, activities, solubilities, etc of systems of furfural, FA, FFA, FDCA, and their sodium salts as functions of absolute pressure and CO_2 partial pressure. These predictions should be applied to evaluate the limiting conversion of the reversible carboxylation reactions of furfural and furoic acid. These predictions may require experimental determination of the equilibrium constant for each reaction under known conditions. The carboxylation control experiments in CHAPTER 3, section 3.3.5.2, Whole-cell and Lysate Biocatalysis, suggest that this data could be collected without active enzyme catalysts, however improved experimental techniques are required to control and monitor bicarbonate concentration and CO_2 pressure. Additionally, improved analytical techniques should be developed to better manage the high salt levels that tend to disrupt HPLC analysis.

Successful characterization and modelling of sodium-furfural-FA-FFA-FDCA-water-CO_2-O_2 systems will be necessary to support reaction engineering and process development tasks, as discussed in the next section.

6.2.3 Future Work in Reaction Engineering and Process Development

The initial goal of this thesis, to develop an enzymatic route from furfural and CO_2 to FDCA presents interesting and unique opportunities and design constraints for reaction engineering and process development that are not available for exploration in other common reaction systems.

The use of enzymes as catalysts effectively limits the reactive phase to aqueous solution. [j] The dense organic liquid phase (Figure 25) may be available for adding furfural into solution, the gas phase may be available for control of dissolved oxygen and carbon dioxide, and the solid phase may be available for product removal.

All three species that may be added to the reaction system in separate fluid phases (i.e., furfural, oxygen, and carbon dioxide) may have transport limitations that should be evaluated in reactor design. It is already commonplace to consider the oxygen mass transfer coefficient, $K_L a$, in bioreactor design, so existing strategies may be applied here. [258-259] Additionally, care should be taken to ensure that sparging oxygen into the reactor does not have the unintended effect of stripping furfural out of it. These considerations may be assisted by vapor pressure modelling (Figure 26).

A variety of alternatives for continuous manufacturing might be considered, however, depending on overall conversion and reactor configuration. Some of these options are introduced below, and the development and evaluation of each alternative might be considered a reasonable research project at some future point in time.

[j] It may also be possible to conduct enzymatic furfural carboxylation directly in supercritical CO_2, but that would require a great deal more research and development, and is beyond the scope of this discussion. [254-257]

6.2.3.1 CSTR without Recycle

As discussed in CHAPTER 3, bicarbonate is believed to be the preferred inorganic carbon source for HmfF and other reversible decarboxylases. If an enzyme membrane reactor or CSTR with immobilized enzyme is used for both carboxylation and oxidation in the same vessel, sodium bicarbonate could be introduced directly into the reactor as a solution or in powdered form. If the inorganic carbon feed is carbon dioxide sourced from direct air capture processes, such as by temperature or vacuum swing adsorption onto amine-functionalized metal organic frameworks (MOF), [260-261] it will have to be converted to bicarbonate before it is accessible for carboxylation. This may be accomplished upstream of the reactor with a sodium hydroxide falling film absorber, or by pressurizing the reactor with carbon dioxide and adding sodium hydroxide at an equimolar rate to dissolution. [262-263] Hydration of carbon dioxide may be accelerated with carbonic anhydrase, as mentioned in section 6.2.1.

Due to the equilibrium between carbon dioxide, carbonic acid, bicarbonate, and carbonate (shown in Figure 53), the inorganic carbon source can perform double-duty as a pH-buffer. pH-adjustments for the reactor may be performed by adding carbonic acid or disodium carbonate – or even possibly by regulating the pressure of the reactor and adding carbonic anhydrase to ensure rapid equilibration between carbonic acid and carbon dioxide. A pH of around 8.5 would ensure that bicarbonate is the dominant inorganic carbon species in solution.

FDCA is most easily and most commonly purified from aqueous solution by acid precipitation with hydrochloric acid. [49, 54] Indeed, the data collected in CHAPTER 4 and discussed in Figure 28 suggest that protonated FDCA is the least soluble of the set FA, FFA, FDCA by at least an order of magnitude, which would facilitate such purification methods. The high buffering capacity of bicarbonate, the high exit pH of the CSTR, and the low pH required to fully protonate FDCA all suggest that a large amount of strong acid equivalents may be required for acid precipitation. It may be reasonable to cool the reactor effluent to enhance FDCA crystallization, although doing so could also have the unintended effect of reducing the difference in supersaturation between protonated FDCA, FFA, and FA (Figure 27), which could cause coprecipitation if either other carboxylic acid is present in the reactor effluent.

6.2.3.2 CSTR with Recycle

If only modest conversion is achieved in a CSTR, then it may be necessary to recycle unreacted aldehyde (possibly FFA, but more likely furfural) back into the reactor to achieve a higher residence time. However, acid precipitation of FDCA would leave any unreacted furfural in an aqueous brine that cannot be directly recycled due to salt accumulation, and the furfural would need to be removed from aqueous solution prior to recycling. Process modelling and technoeconomic studies have been performed on the separation of furfural-water mixtures by distillation. [246, 264] Furfural has a higher boiling point than water (161.7 °C at one atmosphere) [246], so distillation of trace remaining furfural from bulk water would likely be very energy intensive as a great deal of water would have to be boiled off. Furthermore, Curtis (1948) [265] found that sulfate salts of sodium,

magnesium, zinc, and aluminium increase the critical solution temperature of furfural-water mixtures, and that salts readily precipitate from wet furfural (Figure 56). It is likely that similar behavior would occur during the distillation of any reactor effluent, and that solid sodium salts would accumulate in the furfural-rich column bottoms, which might require complicated processing equipment.

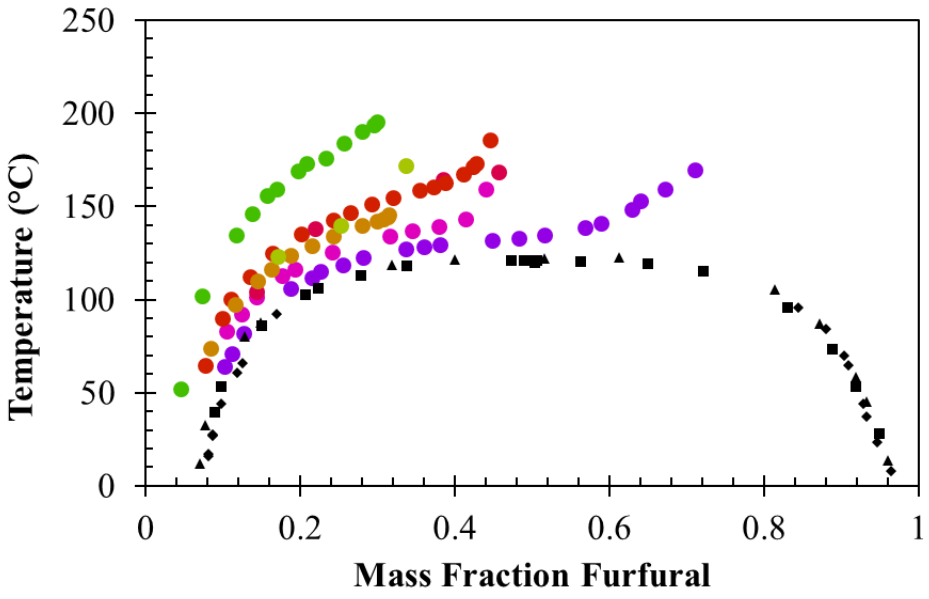


Figure 56: Phase diagram for liquid-liquid equilibrium (LLE) of furfural water systems with and without sulfate salts. Data series plotted in black are the LLE for furfural and water without salt from Rothmund (1898) [138] (▲), Mains (1922) [139] (◆), and Evans and Aylesworth (1926) [140] (■). Data series plotted in colored circles are LLE for furfural and water with a variety of sulfate salts (sodium, aluminum, magnesium, and zinc sulfates) from Curtis (1948) [265]. They are plotted in by color in order of increasing approximate ionic strength: 0.05 (●), 0.08 (●), 0.19 (●), 0.21 (●), 0.22 (●), 0.25 (●), and 0.42 (●).

Although the sulfate data in Figure 56 do not extend to low temperatures, they do hint at the possibility of a more energy efficient separation for recovering small amounts of furfural from salty aqueous solution. It may be possible to cool the effluent from the FDCA crystallizer enough to pass the water-rich spinodal boundary and reconstitute a separate phase of liquid furfural. Pure furfural has a freezing point of -36.5 °C [246], and the dissolved salts (which are unlikely to partition into a furfural phase) will likely contribute some degree of freezing-point depression to the aqueous solution. If these effects are sufficient to enable a cold liquid-liquid phase separation, it may be more energy and cost efficient than any separation process involving an actual phase change of water (i.e., boiling or freezing). Remaining trace furfural could be removed with organic solvent extraction using benzene or toluene, and recovered by distillation (this process could also be done as an alternative to the cold-separation altogether, but higher flux through distillation may prove to be more expensive). [266] Brine could then be optionally treated by reverse osmosis or ion exchange prior to discharging.

6.2.3.3 Batch or Plug Flow Reactor with Recycle

The work in CHAPTER 5, and specifically in section 5.2, A High-Throughput pH-Based Colorimetric Assay: Application Focus on Alpha/Beta Hydrolases focused on tracking and modelling the pH changes that occur over the course of reactions that consume or produce acids and bases. These lessons are of course extensible to the conversion of furfural to FDCA as well, and may facilitate the design of more efficient reactor schemes. As mentioned previously, the high buffering capacity of bicarbonate must be overcome to precipitate FDCA from the effluent of a CSTR operating at a basic pH. However, if a plug

flow reactor with immobilized enzyme is used, some amount of the buffer itself will be consumed, and the pH will fall along the length of the reactor as carboxylic acids are produced.

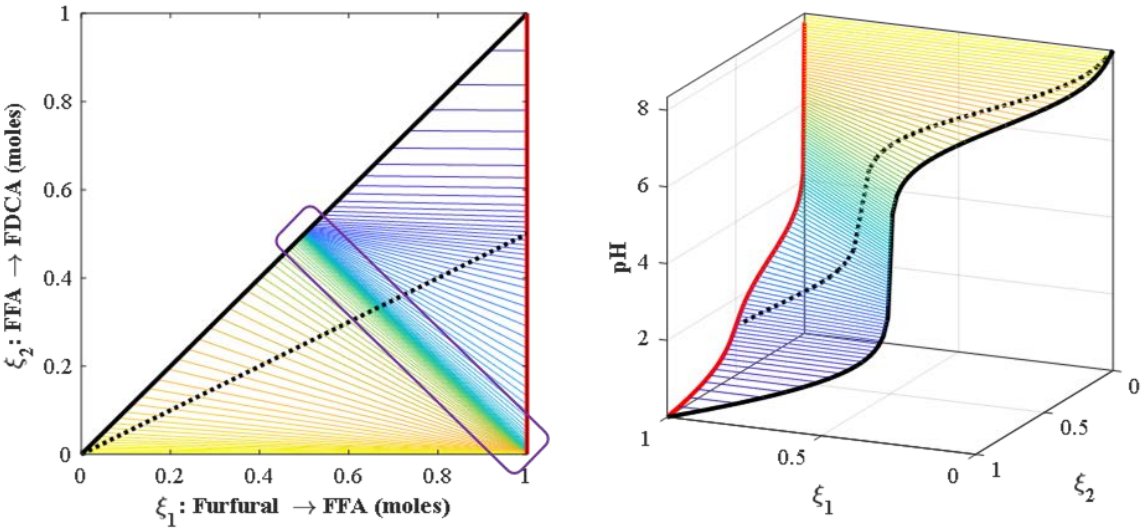


Figure 57: Hypothetical carboxylation-oxidation-pH space of a batch or plug flow reactor converting furfural to FDCA. These plots show the approximate pH (colored isoclines, ▨) of a solution initially containing 1 M sodium bicarbonate as a function of the molar extents of reaction of furfural to FFA (ξ_1) and of FFA to FDCA (ξ_2). Bicarbonate is consumed to produce FFA, and FFA is consumed to produce FDCA, limiting the feasible space to that in which $\xi_2 \leq \xi_1 \leq [NaHCO_3]_{initial}$, as illustrated by the solid red (▬) and black (▬) bounding curves. The accessible reaction space will be further constrained by the pH-stability of the biocatalysts, the solubility limits of all species, specific interactions and activity effects, and (if a gas phase is present) the vapor pressure of carbon dioxide. These features have not been considered in this model, which is for illustrative purposes only. The purple box (◇) on the left-hand plot encompasses the region of greatest slope, in which the production of the second carboxylic acid group on FDCA causes the number of carboxylate equivalents to surpass the amount of sodium cation in solution. The dashed black line (...) traverses an arbitrarily selected path through the extent-coordinates (one of an infinite number of possible paths).

The rough calculations used to produce Figure 57 reveal that the series reactions of furfural and FFA (Figure 8) define a two-dimensional envelope of molar reaction extents ξ_1, and ξ_2 for carboxylation and oxidation, respectively. Using an initial 1 M of sodium bicarbonate as both reactant and buffer distorts this envelope into the pH-dimension, which is parametric in ξ_1 and ξ_2. Any future reaction engineering for batch or plug flow reactors can be thought of as evaluating routes along this warped sheet (or similar). An example of a hypothetical route is displayed in Figure 57, but the specific path traversed by a reactor will depending on the relative rates of each reaction which will be influenced by biocatalyst kinetics and loading. Due to the simplistic nature of the model, it shows only the extreme bounds of what is possible using the mathematics of pH only. The envelope of space that is actually feasible or practical is likely to be much more tightly constrained.

This exercise reveals that certain reactor configurations may offer product streams at lower pH values and buffer capacities than in the CSTR case. With small amounts of additional buffers, it may be possible define a plug flow reactor and an operational regime that maintains a relatively level pH, but depletes alkaline buffer capacity. In this way, it may be possible to keep biocatalysts stable and facilitate a high degree of conversion, but substantially reduce the amount of acid that would be required for FDCA precipitation. This would in turn reduce the salt content that must be contended with after the crystallizer, or prior to recycling any unreacted aldehyde materials.

Additionally, this design approach could be used to generate pH “zones” within a plug flow reactor, which may be useful if it turns out that HMFO or catalase have better

activity or operational stability in more neutral solutions than are required to drive carboxylation.

It is tempting to imagine that reactor effluent could be pressurized with carbon dioxide to acidify the solution and cause FDCA to precipitate without adding any additional salts. Unfortunately, carbonic acid is much less acidic than FDCA, and a similar exercise to that in Figure 57 reveals that the addition of carbon dioxide to an FDCA solution has no effect on pH in the range required to protonate FDCA (although there may be some minor ion and activity-related effects).

APPENDIX A. SEQUENCE INFORMATION FOR BIOCATALYSTS

This appendix contains protein and DNA sequences for the proteins evaluated in this study.

A.1 Protein Amino Acid Sequences

The primary structures of each protein are provided below.

A.1.1 2,3-Dihydroxybenzoic Acid Decarboxylase (2,3-DHBD_Ao)

```
MLGKIALEEAFALPRFEEKTRWWASLFSTDAETHVKEITDINKIRIEHADKHGVGYQILSYTAPG
VQDIWDPVEAQALAVEINDYIAEQVRVNPDRFGAFATLSMHNPKEAADELRRCVEKYGFKGALVN
DTQRAGPDGDDMIFYDNADWDIFWQTCTELDVPFYMHPRNPTGTIYEKLWADRKWLVGPPLSFAH
GVSLHVLGMVTNGVFDRHPKLQIIMGHLGEHVPFDMWRINHWFEDRKKLLGLAETCKKTIRDYFA
ENIWITTSGHFSTTTLNFCMAEVGSDRILFSIDYPFETFSDACEWFDNAELNGTDRLKIGRENAK
KLFKLDSYKDSSA
```

A.1.2 2,6-Dihydroxybenzoic Acid Decarboxylase (2,6-DHBD_Rs)

```
MQGKVALEEHFAIPETLQDSAGFVPGDYWKELQHRLLDIQDTRLKLMDAHGIETMILSLNAPAVQ
AIPDRRKAIEIARRANDVLAEECAKRPDRFLAFAALPLQDPDAATEELQRCVNDLGFVGALVNGF
SQEGDGQTPLYYDLPQYRPFWGEVEKLDVPFYLHPRNPLPQDSRIYDGHPWLLGPTWAFAQETAV
HALRLMASGLFDEHPRLNIILGHMGEGLPYMMWRIDHRNAWVKLPPRYPAKRRFMDYFNENFHIT
TSGNFRTQTLIDAILEIGADRILFSTDWPFENIDHASDWFNATSIAEADRVKIGRTNARRLFKLD
GA
```

A.1.3 Salicylic Acid Decarboxylase (SAD_Tm)

```
MRGKVSLEEAFELPKFAAQTKEKAELYIAPNNRDRYFEEILNPCGNRLELSNKHGIGYTIYSIYS
PGPQGWTERAECEEYARECNDYISGEIANHKDRMGAFAALSMHDPKQASEELTRCVKELGFLGAL
VNDVQHAGPEGETHIFYDQPEWDIFWQTCVDLDVPFYLHPEPPFGSYLRNQYEGRKYLIGPPVSF
ANGVSLHVLGMIVNGVFDRFPKLKVILGHLGEHIPGDFWRIEHWFEHCSRPLAKSRGDVFAEKPL
LHYFRNNIWLTTSGNFSTETLKFCVEHVGAERILFSVDSPYEHIDVGCGWYDDNAKAIMEAVGGE
KAYKDIGRDNAKKLFKLGKFYDSEA
```

A.1.4 2,5-Furancarboxylic Acid Decarboxylase 1 (HmfF)

```
[M]{G}SRQPAEKTNSASQAAPLPEGPLTLRSWLRHLGNTDRLAAIDEPVALEHTLAAVAKRLDG
ERAVLFRRPGGHAVPVVSGFMSRRAWIAEAMGVPEAGLLERMRSAAAQPLPVSEVAQGEAACQQV
IHLDKVDLHKLLPIPTHSEHDNGPYITAGLAIARNPRTGVQNVSIHRIQVHAADRMAILLLPRHL
DAFYRAAEECGEALPIAIVIGVDPLTMLASQAITPIDYDELEIAGALHGAPLEVIKCRTSDVRVP
ANAEIVIEGRLLPGEREMEGPFGEFPKYYSSAEPREVIQVDAVTHRHRPIYHTIVPAEMEHLLLG
AIPREATLLAHLQRSHPGVQDVHLSVGGVCRYHLYVKLDKKREGEAKNVILSAFGAHYDIKQVVV
VDTDVDVHDPAEVEWAVATRFQADQDLVVIAGAQGSVLDPSTTVAANLAGIDNPEPHLQGICAKM
GLDATRPVKYAAHVFTRVRIPGESTIDLQALVSVDPSQWEAYLGEGA
```

[] = Residue removed during construction of fusion proteins.
{} = Residue added in pETDuet expression only.

A.1.5 Flavin Prenyltransferase UbiX (HmfG)

```
[M]CSRACGFPASQPSICRRWCRSTHRNGKRISAKELDVMASQRRIIVGISGASGAAIGVNLLKA
MRGLDGVESHLIVSASGMLTATQELGIKRSELEALADVVHNVRDIGAAVASGSFVTEGMVVAPCS
MKTLASVANGFSDNLLTRAADVVLKERRRLVLVARETPLNLAHLRNMLHATEMGAIVMPPVPAFY
SHPTSIEDVVNHTVGRILDLFQIEHGTLVSRWSGLAHDFAR
```

[] = Residue removed during construction of fusion proteins.

A.1.6 Flavin Prenyltransferase UbiX (UbiX)

```
MSGPERITLAMTGASGAQYGLRLLDCLVQEEREVHFLISKAAQLVMATETDVALPAKPQAMQAFL
TEYCGAAAGQIRVFGQNDWMAPPASGSSAPNAMVICPCSTGTLSAVATGACNNLIERAADVALKE
RRPLVLVPREAPFSSIHLENMLKLSNLGAVILPAAPGFYHQPQSVEDLVDFVVARILNTLGIPQD
MLPRWGEQHLVSDE
```

A.1.7 HMF Oxidase V367R-W466F (HMFO)

```
[M{G}]TDTIFDYVIVGGGTAGSVLANRLSARPENRVLLIEAGIDTPENNIPPEIHDGLRPWLPR
LSGDKFFWPNLTIHRAAEHPGITREPQFYEQGRLLGGGSSVNMVVSNRGLPRDYDEWQALGADGW
DWQGVLPYFIKTERDADYGDDPLHGNAGPIPIGRVDSRHWSDFTVAATQALEAAGLPNIHDQNAR
FDDGYFPPAFTLKGEERFSAARGYLDASVRVRPNLSLWTESRVLKLLTTGNAITGVSVLRGRETL
QVQAREVILTAGALQSPAILLRTGIGPAADLHALGIPVLADRPGVGRNLWEHSSIGVVAPLTEQA
RADASTGKAGSRHQLGIRASSGVDPATPSDLFLHIGADPVSGLASARFWVNKPSSTGWLKLKDAD
PFSYPDVDFNLLSDPRDLGRLKAGLRLITHYFAAPSLAKYGLALALSRFAAPQPGGPLLNDLLQD
EAALERYLRTNVGGVFHASGTARIGRADDSQAVVDKAGRVYGVTGLRVADASIMPTVPTANTNLP
TLMLAEKIADAILTQALQENLYFQGAGWSHPQFEK
```

`[]` = Residues removed during construction of fusion proteins.
`{}` = Residue added, not present in original published protein.

A.1.8 Hexahistidine Affinity Tag and SUMO Solubility Tag

```
MGHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQG
KEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG
```

A.2 Protein DNA Sequences

This appendix section contains gene sequences that were ordered from Genscript (Picastaway, NJ) for the construction of HmfF, HmfG, and HMFO.

A.2.1 2,5-Furancarboxylic Acid Decarboxylase 1 (HmfF)

GCGC(G▾AATTC)[ATG]{GGC}TCGCGCCAGCCGGCCGAAAAAACCAACAGTGCGAGCCAAGCCG
CGCCGCTGCCGGAAGGTCCATTGACGCTGCGCTCATGGTTGCGCCATTTGGGCAACACCGACCGC
TTAGCTGCCATTGATGAGCCTGTGGCGCTGGAGCATACACTGGCAGCGGTCGCAAAACGCCTCGA
CGGCGAGCGCGCGGTGCTGTTCCGTCGCCCGGGTGGCCATGCCGTACCTGTCGTAAGTGGCTTTA
TGAGCCGTCGCGCATGGATCGCGGAGGCAATGGGCGTTCCGGAAGCAGGCCTGTTGGAACGCATG
CGCTCCGCAGCCGCCCAACCGCTGCCGGTGAGCGAGGTCGCCCAGGGCGAAGCGGCATGTCAGCA
GGTGATTCATCTTGATAAAGTCGATCTCCATAAACTTCTGCCTATTCCGACCCATTCGGAACATG
ATAACGGTCCGTATATTACCGCAGGTTTGGCGATCGCACGCAACCCTCGTACGGGTGTTCAGAAC
GTAAGCATCCACCGTATTCAGGTCCATGCTGCCGATCGTATGGCAATTCTGCTGCTGCCGCGCCA
CCTGGATGCGTTTTACCGCGCCGCGGAAGAATGCGGCGAAGCATTACCGATCGCGATCGTGATTG
GTGTGGATCCGTTAACGATGTTAGCTAGCCAAGCGATCACTCCAATCGACTATGATGAGCTGGAG
ATCGCCGGGGCGCTGCATGGAGCTCCACTGGAGGTGATCAAATGTCGCACCTCGGATGTTCGTGT
CCCGGCGAATGCGGAAATTGTGATCGAGGGTCGCCTGCTGCCGGGTGAGCGCGAAATGGAAGGGC
CTTTCGGCGAATTTCCGAAATACTATTCTTCGGCGGAGCCGCGCGAAGTTATTCAGGTTGACGCA
GTTACCCATCGCCACCGCCCAATCTATCATACAATCGTGCCTGCGGAAATGGAACATCTGCTGCT
GGGTGCCATTCCTCGTGAAGCCACTCTGCTGGCCCACCTCCAGCGTAGTCACCCGGGAGTACAAG
ATGTACATCTGAGCGTTGGTGGCGTTTGCCGTTACCACCTGTATGTTAAATTAGATAAAAAACGC
GAGGGAGAGGCCAAAAATGTAATTTTGAGTGCGTTTGGTGCCCATTATGACATCAAGCAGGTTGT
GGTGGTGGATACAGATGTTGATGTACACGATCCGGCGGAAGTTGAATGGGCAGTGGCTACTCGTT
TTCAGGCAGACCAGGATCTGGTTGTGATCGCTGGGGCGCAAGGGTCAGTCCTGGATCCGAGCACC
ACGGTAGCGGCGAATCTGGCTGGGATTGACAATCCAGAACCTCATCTCCAGGGCATTTGTGCGAA
AATGGGGCTGGACGCTACTCGCCCTGTAAAATATGCGGCACATGTGTTTACCCGCGTGCGCATTC
CTGGCGAATCTACAATTGATTTACAGGCTCTGGTCTCTGTTGATCCTTCCCAGTGGGAGGCCTAT
CTGGGAGAAGGCGCATGATAA(A▴AGCTT)GCGC

(▾) = Leading restriction site: EcoRI.
(▴) = Lagging restriction site: HindIII.
[] = Base pairs removed during construction of fusion proteins.
{} = Base pairs added in pETDuet expression only.

A.2.2 Flavin Prenyltransferase UbiX (HmfG)

GCGCGCGC(CA▾T[ATG])TGTTCACGCGCCTGCGGGTTTCCGGCGTCTCAGCCATCGATCTGTCG
CCGTTGGTGCCGCAGCACGCACCGCAATGGCAAGCGCATTAGCGCCAAAGAATTGGATGTGATGG
CTTCTCAGCGCCGTATCATTGTCGGTATCTCGGGCGCCTCTGGAGCGGCGATCGGTGTTAACTTA
TTGAAAGCGATGCGCGGTTTGGACGGTGTTGAGTCTCATCTGATCGTGTCTGCAAGCGGTATGCT
GACGGCTACACAAGAACTGGGCATTAAACGTTCAGAATTGGAAGCGCTGGCAGATGTCGTGCACA
ATGTACGCGACATCGGAGCTGCGGTCGCGAGCGGTAGCTTTGTAACCGAGGGCATGGTGGTGGCC
CCATGCAGCATGAAAACGCTCGCATCGGTGGCAAATGGCTTTTCGGACAACTTACTCACCCGCGC
GGCAGATGTTGTTCTGAAGGAACGTCGTCGTCTTGTGTTGGTTGCCCGCGAAACTCCGCTCAATC
TTGCCCATTTACGCAATATGCTTCACGCAACGGAAATGGGCGCGATCGTGATGCCGCCTGTCCCG
GCCTTTTACAGTCATCCGACGAGCATTGAGGACGTTGTGAACCACACTGTAGGGCGCATCCTGGA
TCTTTTTCAAATTGAGCATGGGACACTTGTGAGCCGTTGGTCTGGACTGGCACACGATTTCGCCC
GTTGATAA(CTC▴GAG)GCGC

(▾) = Leading restriction site: NdeI.
(▴) = Lagging restriction site: XhoI.
[] = Base pairs removed during construction of fusion proteins.

A.2.3 HMF Oxidase V367R-W466F (HMFO)

GCGCGCGCG(CA▾[TATG){GGG}]ACTGATACGATTTTTGACTACGTGATTGTTGGCGGTGGCAC
GGCGGGTAGCGTTCTGGCCAACCGTCTGTCCGCCCGTCCGGAGAATCGCGTGTTGCTGATTGAGG
CCGGTATTGATACCCCGGAAAACAATATTCCGCCGGAGATCCACGATGGCCTGCGCCCGTGGCTG
CCGCGTCTGAGCGGTGATAAGTTCTTTTGGCCGAATCTGACCATCCACCGTGCCGCGGAACACCC
GGGTATCACGCGCGAGCCGCAGTTCTATGAACAAGGCCGTCTGCTGGGCGGTGGTAGCAGCGTGA
ACATGGTCGTTTCTAACCGTGGTCTGCCTCGCGACTATGACGAATGGCAGGCACTGGGCGCAGAT
GGTTGGGATTGGCAGGGTGTTCTGCCGTACTTCATCAAGACCGAGCGTGACGCGGACTACGGTGA
CGACCCGTTGCATGGCAATGCGGGTCCGATTCCGATCGGTCGCGTCGATTCGCGTCACTGGAGCG
ACTTCACGGTGGCGGCAACCCAAGCTCTGGAAGCGGCTGGCCTGCCGAACATTCACGACCAAAAC
GCACGTTTTGATGACGGTTACTTCCCACCGGCATTTACGTTGAAAGGTGAAGAGCGCTTCAGCGC
CGCACGCGGTTATCTGGATGCGAGCGTCCGTGTGCGTCCGAACCTGAGCCTGTGGACTGAGAGCC
GTGTCCTGAAGCTGCTGACCACTGGCAATGCAATCACCGGTGTGAGCGTGCTGCGTGGTCGCGAA
ACCCTGCAAGTTCAAGCGCGCGAGGTCATCCTGACCGCCGGTGCGTTGCAAAGCCCAGCGATTCT
GTTGCGCACCGGCATCGGCCCTGCGGCGGATCTGCACGCACTGGGTATTCCTGTTCTGGCAGACC
GTCCGGGTGTTGGTCGCAATCTGTGGGAGCACAGCTCTATCGGTGTGGTTGCCCCGCTGACCGAG
CAGGCACGTGCAGACGCCAGCACGGGTAAAGCCGGCTCTCGCCATCAACTGGGTATCCGTGCGTC
GTCCGGCGTAGATCCGGCGACGCCTAGCGACCTGTTTCTGCATATCGGTGCTGATCCAGTCAGCG
GTCTGGCAAGCGCTCGTTTCTGGGTGAACAAGCCAAGCTCCACCGGCTGGCTGAAGCTGAAGGAC
GCGGACCCGTTTAGCTACCCGGACGTAGACTTCAATCTGCTGAGCGATCCGCGCGACTTGGGTCG
TCTGAAAGCGGGCCTGCGTCTGATCACCCATTACTTCGCAGCGCCGTCCCTGGCGAAATATGGTT
TGGCGCTGGCATTGAGCCGTTTTGCGGCACCGCAGCCGGGTGGTCCGCTGCTGAACGACCTGTTG
CAGGACGAAGCCGCCCTGGAACGCTATTTGCGTACGAACGTCGGCGGTGTTTTCCATGCGAGCGG
CACGGCGCGTATCGGCCGTGCGGATGATTCCCAGGCTGTTGTCGATAAAGCGGGTCGTGTGTACG
GCGTCACCGGCCTGCGTGTTGCGGACGCAAGCATTATGCCGACCGTTCCGACCGCCAATACCAAT
CTGCCGACGCTGATGCTGGCTGAGAAAATTGCGGATGCGATTCTGACCCAGGCTCTGCAGGAGAA
TTTATATTTTCAAGGTGCCGGCTGGAGCCACCCGCAGTTTGAAAAATAA(C▴TCGAG)GCGCGC

(▾) = Leading restriction site: NdeI.
(▴) = Lagging restriction site: XhoI.
[] = Base pairs removed during construction of fusion proteins.
{} = Codon added, not present in original published protein.

A.2.4 Hexahistidine Affinity Tag and SUMO Solubility Tag

…AGATATA(C▾CATGG)GTCATCACCATCATCATCACGGGTCGGACTCAGAAGTCAATCAAGAAGC
TAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGAT
CTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCT
AAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTTGTACGACGGTATTAGAATTCAAGCTGA
TCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACAGAGAACAGATTG
GTGGT[CAT…]

(▾) = Leading restriction site: NcoI.
[] = base pair(s) removed during construction of fusion proteins.

A.3 Primer Sequences

The primer sequences used for PCR in this work are listed below. The use of each primer is described in section 3.3.4.1 Cloning of Enzymes for Heterologous Expression, and in Table 7.

A: CCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGG
B: CCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTG
1: GCGCGC(CA▾TATG}TCGCGCCAGCCGGCCGAAAAAAC
2: GCGC(C˅CATGG)GCTCGCGCCAGCCGGCCGAAAAAAC
3: GCGC(AA▴GCTT)TTATCATGCGCCTTCTCCCAG
4: ACAGAGAACAGATTGGTGGTTCGCGCCAGCCGGCCG
5: CGGCCGGCTGGCGCGAACCACCAATCTGTTCTCTGT
6: ACAGAGAACAGATTGGTGGTTGTTCACGCGCCTGC
7: GCAGGCGCGTGAACAACCACCAATCTGTTCTCTGT
8: ACAGAGAACAGATTGGTGGTACTGATACGATTTTTGACTACGTG
9: CACGTAGTCAAAAATCGTATCAGTACCACCAATCTGTTCTCTGT

(▾) = Leading restriction site NdeI.
(˅) = Leading Restriction site NcoI.
(▴) = Lagging restriction site HindIII.

APPENDIX B. TABULATED PHYSICAL PROPERTY DATA FOR FURFURAL, FA, FFA, AND FDCA

This appendix contains the solubility data that were collected for furfural, FA, FFA, and FDCA for PC-SAFT modeling as described in Section 4.2, Mutual Influence of Furfural and Furancarboxylic Acids on Their Solubility in Aqueous Solutions: Experiments and PC-SAFT Predictions.

B.1 Single-Solute Data

Table 21: Solubility of furfural and the furancarboxylic acids FA, FFA, and FDCA in pure solvent and in ethanol/water solvent mixtures between 18 °C and 45 °C at atmospheric pressure as well as pH of saturated solutions at 35 °C and atmospheric pressure in water without ethanol. Standard uncertainties u are u(T) = 0.1 K, u(pH) = 0.05, u(p) = 0.2 bar and u(w) = 0.001 g solute/(g total).

Solute	T °C	Mass Fraction of Water g water / (g water + ethanol)	pH $-\log_{10}(a_H)$	Solubility g / (g total)
Furfural	35	1	-	1.02E-01
Furfural	35	1	2.56	-
FA	35	1	-	6.86E-02
FA	35	1	1.77	-
FFA	18	0	-	5.49E-02
FFA	25	0	-	6.25E-02
FFA	32	0	-	7.12E-02
FFA	37	0	-	7.77E-02
FFA	45	0	-	8.99E-02
FFA	35	0.25	-	9.16E-02
FFA	35	0.5	-	5.87E-02
FFA	35	0.75	-	2.09E-02
FFA	18	1	-	5.70E-03
FFA	25	1	-	7.32E-03
FFA	32	1	-	8.23E-03
FFA	35	1	-	1.19E-02
FFA	35	1	1.76	-

Table 21 (continued)

Solute	T °C	Mass Fraction of Water g water / (g water + ethanol)	pH $-\log_{10}(a_H)$	Solubility g / (g total)
FFA	37	1	-	9.38E-03
FFA	45	1	-	1.02E-02
FDCA	25	0	-	8.89E-03
FDCA	32	0	-	1.05E-02
FDCA	37	0	-	1.13E-02
FDCA	45	0	-	1.44E-02
FDCA	35	0.25	-	1.67E-02
FDCA	35	0.5	-	1.13E-02
FDCA	35	0.75	-	3.35E-03
FDCA	18	1	-	6.28E-04
FDCA	25	1	-	6.99E-04
FDCA	32	1	-	1.05E-03
FDCA	35	1	-	1.46E-03
FDCA	35	1	2.07	-
FDCA	37	1	-	1.15E-03
FDCA	45	1	-	1.41E-03

Table 22: pH-dependent solubility of the furancarboxylic acids FA, FFA, and FDCA in water at 35 °C at atmospheric pressure. Standard uncertainties u are u(T) = 0.1 K, u(pH) = 0.05, u(p) = 0.1 bar and u(w) = 0.001 g solute/(g total).

Solute	pH $-\log_{10}(a_H)$	Solubility g / (g total)	Solute	pH $-\log_{10}(a_H)$	Solubility g / (g total)
FA	1.98	7.19E-02	FFA	1.82	1.01E-02
FA	2.08	8.51E-02	FFA	2.84	4.72E-02
FA	2.23	7.03E-02	FFA	3.08	7.48E-02
FA	2.37	7.77E-02	FFA	3.18	1.31E-01
FA	2.47	8.44E-02	FFA	3.20	1.10E-01
FA	2.56	8.89E-02	FFA	3.21	1.24E-01
FA	2.63	9.09E-02	FFA	3.22	1.11E-01
FA	2.70	9.45E-02	FFA	3.22	1.27E-01
FA	2.76	1.00E-01	FFA	3.22	1.25E-01
FA	2.81	1.02E-01	FFA	3.24	1.41E-01
FA	4.80	3.97E-01	FFA	3.26	1.40E-01
FA	4.82	4.07E-01	FFA	3.26	1.26E-01

Table 22 (continued)

Solute	pH $-\log_{10}(a_H)$	Solubility g / (g total)	Solute	pH $-\log_{10}(a_H)$	Solubility g / (g total)
FA	4.83	3.99E-01	FFA	3.27	1.28E-01
FA	4.84	4.02E-01	FFA	3.27	1.39E-01
FA	4.84	4.07E-01	FFA	3.28	1.25E-01
FA	4.85	4.10E-01	FFA	3.32	1.47E-01
FA	4.86	4.13E-01	FFA	3.38	1.63E-01
FA	4.86	4.12E-01	FFA	4.01	2.11E-01
FA	4.87	4.11E-01	FFA	4.03	2.66E-01
FA	4.88	4.14E-01	FFA	4.06	2.22E-01
FA	4.96	3.96E-01	FFA	4.09	2.63E-01
FA	5.10	3.72E-01	FFA	4.29	1.70E-01
FA	5.87	3.75E-01	FFA	4.35	1.60E-01
FA	6.36	3.73E-01	FFA	4.84	1.55E-01
FA	6.92	3.83E-01	FFA	4.93	1.28E-01
FA	7.38	3.68E-01	FDCA	4.36	6.66E-02
FDCA	2.17	1.34E-03	FDCA	4.39	8.50E-02
FDCA	2.72	4.17E-03	FDCA	4.45	8.30E-02
FDCA	2.92	7.31E-03	FDCA	4.47	9.24E-02
FDCA	3.08	1.16E-02	FDCA	4.50	9.33E-02
FDCA	3.09	1.22E-02	FDCA	4.50	7.47E-02
FDCA	3.16	1.48E-02	FDCA	4.50	9.41E-02
FDCA	3.19	1.64E-02	FDCA	4.54	8.69E-02
FDCA	3.32	2.42E-02	FDCA	4.54	8.40E-02
FDCA	3.33	2.58E-02	FDCA	4.56	9.23E-02
FDCA	3.35	2.71E-02	FDCA	4.57	9.59E-02
FDCA	3.51	3.98E-02	FDCA	4.60	1.00E-01
FDCA	3.53	4.30E-02	FDCA	4.63	1.06E-01
FDCA	3.90	5.15E-02	FDCA	4.64	8.14E-02
FDCA	3.90	4.94E-02	FDCA	4.67	9.84E-02
FDCA	4.10	6.22E-02	FDCA	4.68	8.53E-02
FDCA	4.12	6.57E-02	FDCA	4.72	1.16E-01
FDCA	4.15	6.52E-02	FDCA	4.82	1.08E-01
FDCA	4.16	6.54E-02	FDCA	4.94	1.37E-01
FDCA	4.16	7.07E-02	FDCA	4.99	1.20E-01
FDCA	4.19	6.54E-02	FDCA	5.09	1.53E-01
FDCA	4.30	7.81E-02	FDCA	5.16	1.47E-01
FDCA	4.34	7.46E-02	FDCA	5.28	1.62E-01
FDCA	4.34	7.25E-02	FDCA	5.30	1.81E-01
FDCA	4.34	7.37E-02	FDCA	5.38	1.96E-01

B.2 Mixed-Solute Data

Table 23: Solubility of furfural and the furancarboxylic acids FA, FFA, and FDCA in water at 35 °C in the presence of one of the cosolutes furfural, FA, FFA, or FDCA at atmospheric pressure. Standard uncertainties u are u(T) = 0.1 K, u(pH) = 0.05, u(p) = 0.1 bar and u(w) = 0.001 g solute/(g total).

Primary Solute	Cosolute	Solubility of Primary Solute g solute / (g total)	Cosolute Weight Fraction g cosolute / (g total)
Furfural	FA	1.02E-01	1.63E-02
Furfural	FA	1.02E-01	3.18E-02
Furfural	FA	1.01E-01	3.06E-02
Furfural	FFA	1.02E-01	1.98E-03
Furfural	FFA	1.03E-01	3.52E-03
Furfural	FFA	1.02E-01	4.91E-03
Furfural	FDCA	1.01E-01	2.17E-04
Furfural	FDCA	1.03E-01	4.38E-04
Furfural	FDCA	1.03E-01	9.78E-04
FA	Furfural	1.02E-01	2.82E-02
FA	Furfural	1.34E-01	5.07E-02
FA	Furfural	1.56E-01	8.56E-02
FA	FFA	7.43E-02	2.31E-03
FA	FFA	7.15E-02	5.76E-03
FA	FFA	7.36E-02	8.31E-03
FA	FDCA	7.21E-02	3.78E-04
FA	FDCA	6.88E-02	6.43E-04
FA	FDCA	6.91E-02	1.23E-03
FFA	Furfural	2.08E-02	3.35E-02
FFA	Furfural	2.16E-02	6.30E-02
FFA	Furfural	2.73E-02	9.20E-02
FFA	FA	1.44E-02	2.72E-02
FFA	FA	1.78E-02	5.15E-02
FFA	FA	1.74E-02	5.04E-02
FFA	FDCA	1.17E-02	3.46E-04
FFA	FDCA	1.16E-02	6.35E-04
FFA	FDCA	1.12E-02	6.65E-04
FDCA	Furfural	2.18E-03	3.63E-02

Table 23 (continued)

Primary Solute	Cosolute	Solubility of Primary Solute g solute / (g total)	Cosolute Weight Fraction g cosolute / (g total)
FDCA	Furfural	2.60E-03	6.78E-02
FDCA	Furfural	3.04E-03	9.66E-02
FDCA	FA	1.69E-03	2.80E-02
FDCA	FA	2.01E-03	5.26E-02
FDCA	FA	2.02E-03	5.25E-02
FDCA	FFA	1.16E-03	2.60E-03
FDCA	FFA	8.69E-04	6.26E-03
FDCA	FFA	6.97E-04	9.01E-03

Table 24: Measured pH values of solute-saturated solutions in the presence of a cosolute of known weight fraction at 35 °C and 1 bar. Standard uncertainties u are u(T) = 0.1 K, u(pH) = 0.05, u(p) = 0.1 bar and u(w) = 0.001 g solute/(g total).

Primary Solute	Cosolute	Cosolute Weight Fraction g cosolute / (g total)	pH $-\log_{10}(a_H)$
Furfural	FA	2.45E-02	1.97
Furfural	FA	3.06E-02	1.84
Furfural	FFA	2.79E-03	2.00
Furfural	FFA	4.91E-03	1.84
Furfural	FDCA	3.31E-04	2.08
Furfural	FDCA	9.78E-04	2.03
FA	Furfural	4.09E-02	1.68
FA	Furfural	8.56E-02	1.61
FA	FFA	4.11E-03	1.65
FA	FFA	8.31E-03	1.60
FA	FDCA	5.15E-04	1.67
FA	FDCA	1.23E-03	1.66
FFA	Furfural	5.02E-02	1.67
FFA	Furfural	9.20E-02	1.62
FFA	FA	4.01E-02	1.63
FFA	FA	5.04E-02	1.58
FFA	FDCA	4.95E-04	1.69
FFA	FDCA	6.65E-04	1.68
FDCA	Furfural	5.41E-02	1.98

Table 24 (continued)

Primary Solute	Cosolute	Cosolute Weight Fraction g cosolute / (g total)	pH $-\log_{10}(a_H)$
FDCA	Furfural	9.66E-02	1.92
FDCA	FA	4.10E-02	1.80
FDCA	FA	5.25E-02	1.69
FDCA	FFA	4.51E-03	1.84
FDCA	FFA	9.01E-03	1.74

APPENDIX C. SUPPORTING INFORMATION FOR PH-BASED COLORIMETRIC ASSAY

This section contains additional experimental data in support of and supplement to the text of CHAPTER 5, section 5.2: A High-Throughput pH-Based Colorimetric Assay: Application Focus on Alpha/Beta Hydrolases.

C.1 General supporting information

C.1.1 Workflow and Dependency of Steps

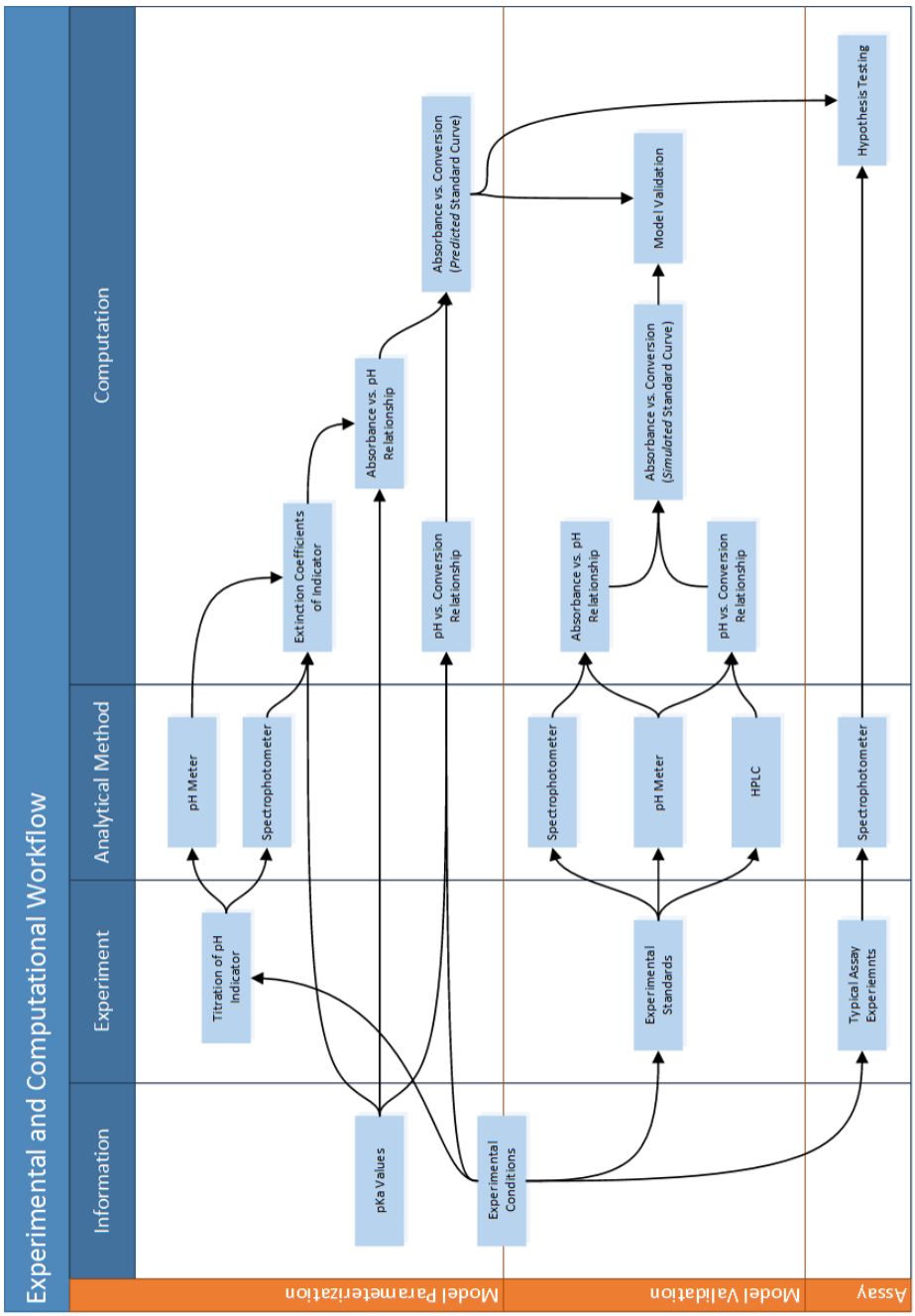


Figure 58: Diagram illustrating the dependencies between steps performed in this work. The model must be parameterized and validated before it can be used for high throughput screening. The experimental conditions at which assays will typically be run are used as a basis for evaluating the colorimetric pH indicator. The assay can be validated using true or simulated standards, which may be checked with various experimental methods. Ultimately, assay experiments are compared to model predictions for hypothesis testing.

C.1.2 Other Assays for Alpha-Beta Hydrolases

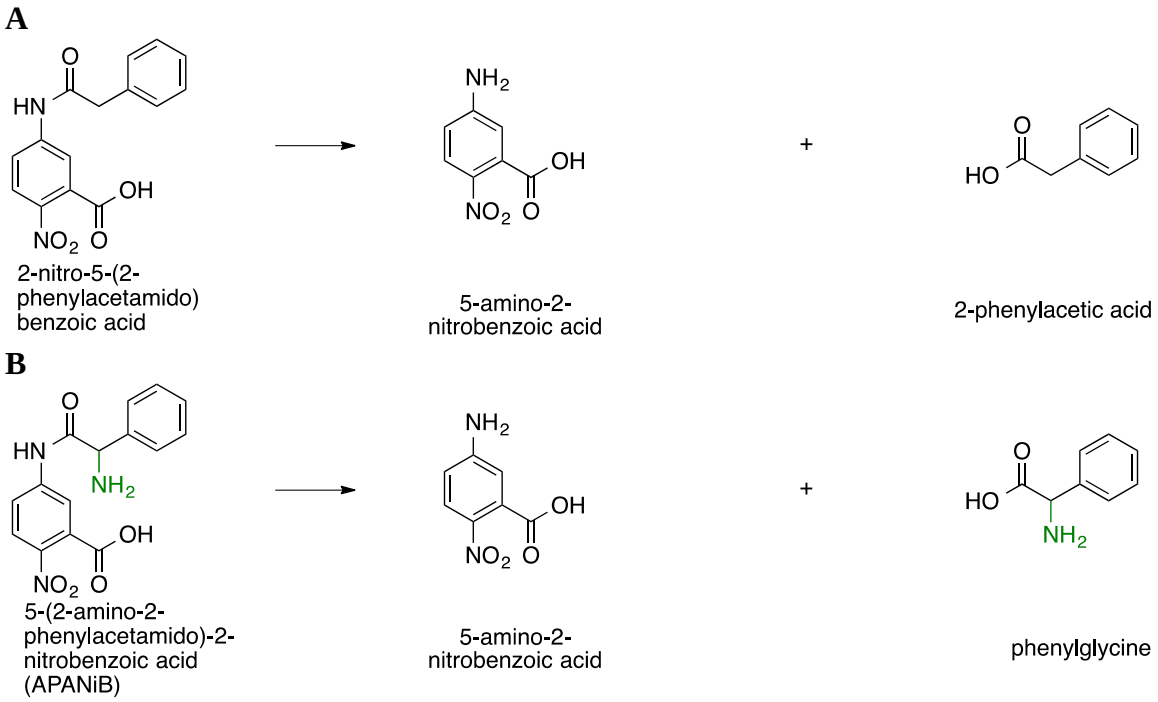


Figure 59: Alternative colorimetric assays for alpha-beta hydrolases. Colorimetric reactions for penicillin G acylase (PGA) (A) and α-amino acid ester hydrolase (AEH) (B) whose hydrolysis yield 405 nm- absorbing compound: 5-nitro-2-amino benzoic acid. (Continued on next page)

C

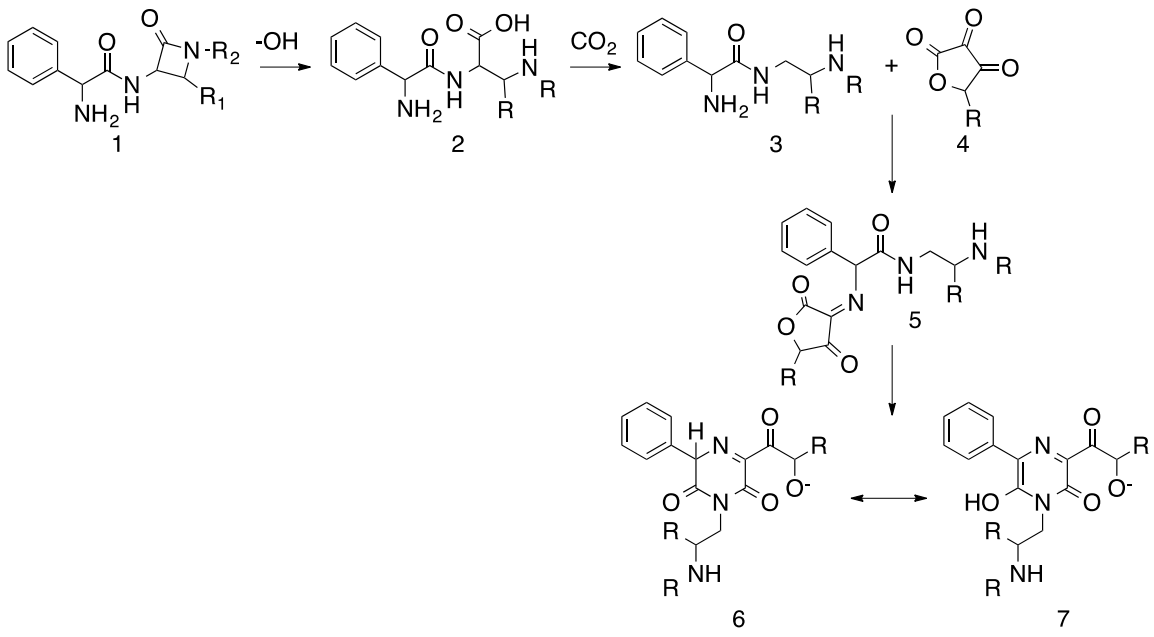


Figure 59 continued. (C) Diketopiperazine reaction that results in a chromogenic compound that absorbs between 406-410 nm. (Continued on next page)

D

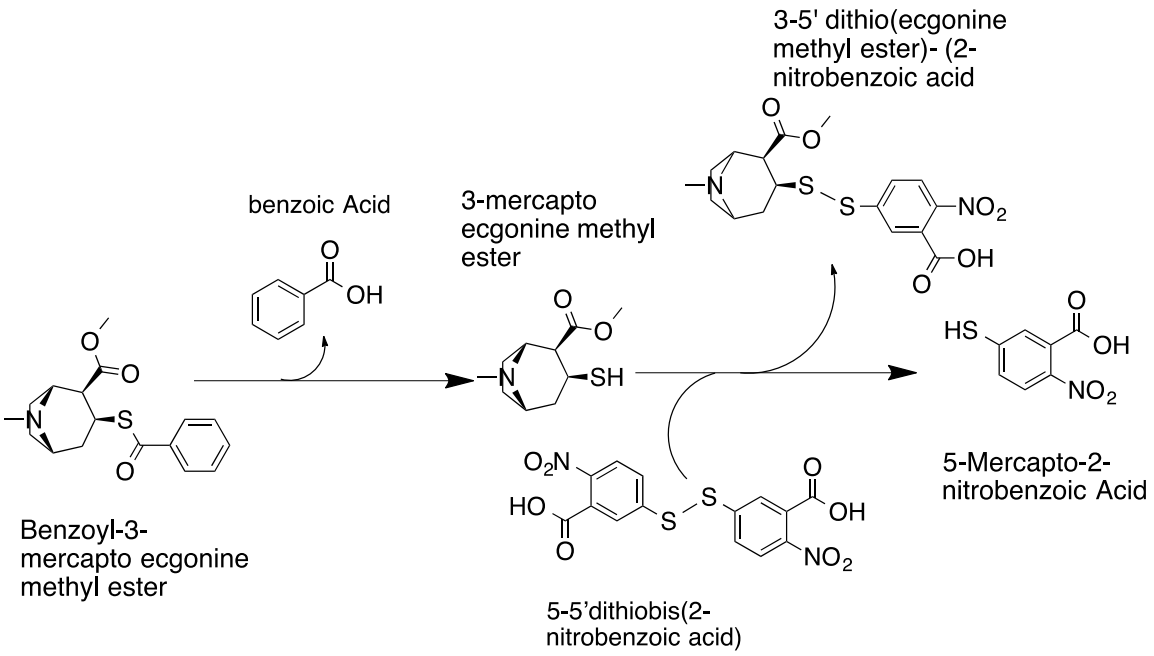


Figure 59 continued. (D) Thioesterase assay for cocaine esterase.[184] In lieu of the native substrate, cocaine, benzoyl 3-mercapto-ecgonine methyl ester is presented to the enzyme. The enzymatic product, 3-mercapto-ecgonine methyl ester, reacts with 5-5'dithionis(2-nitrobenzoic acid) (DNTB) to form the chromophore 5-mercapto-2-nitrobenzoic acid, which is monitored at 412 nm. As neither the target substrate nor target reaction is tested, the results and conclusion from these assays may not be descriptive of the enzymes' activities in the truly desired reaction conditions.

C.1.3 Simulated Conversion Data and Validation of pH for Tracking Reactions

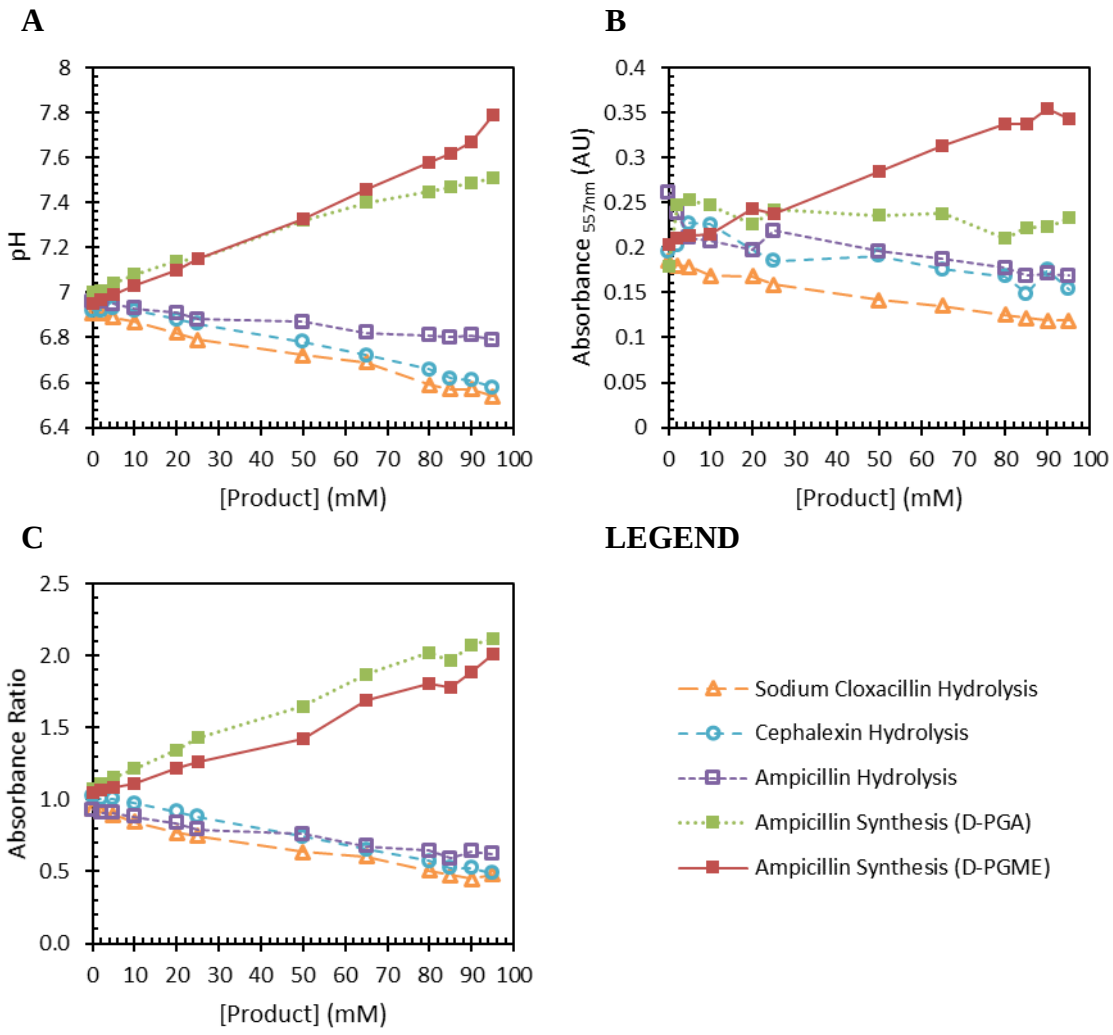


Figure 60: pH and absorbance measurements of samples simulating hydrolysis and synthesis reactions. These data are from non-enzymatic samples that simulate the hydrolysis (open markers) of cloxacillin (diamonds), cephalexin (circles), and ampicillin (squares), as well as the synthesis of ampicillin with D-PGA (dashed green) or with D-PGME (solid red). Sample preparation is described in the Materials & Methods section of the main text. These simulated hydrolysis and synthesis reactions result measurable changes in pH, providing analytical standards that compress the many variables involved in tracking the course of each reaction into a single parameter that is easily evaluated with potentiometry or colorimetry.

C.1.4 Validation of Phenol Red

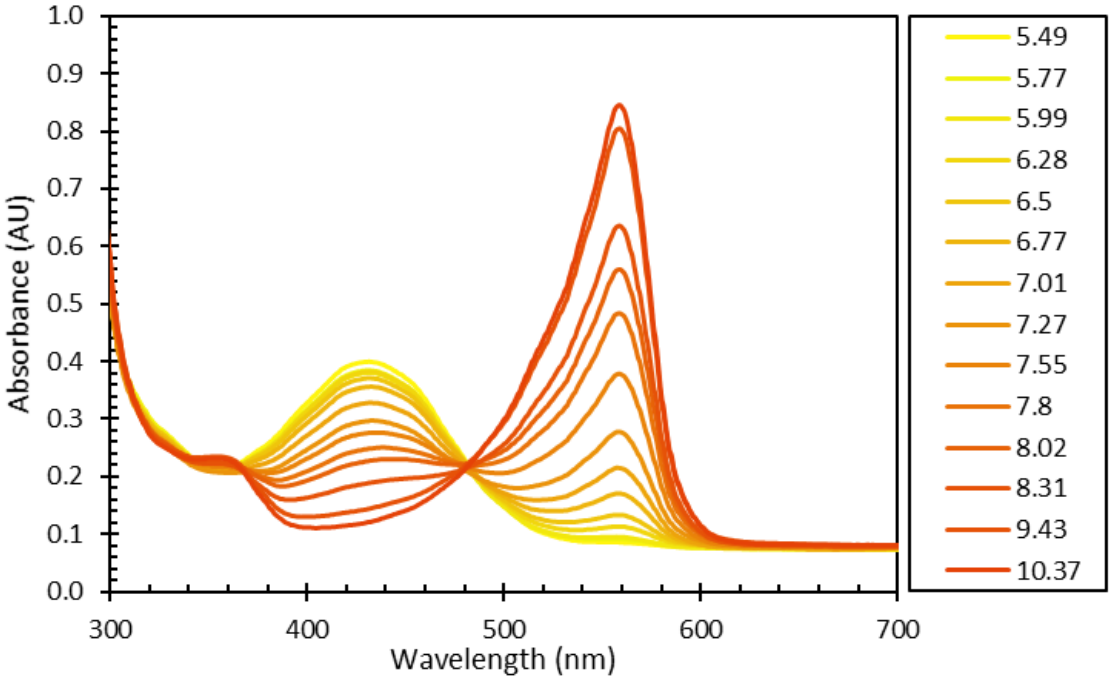
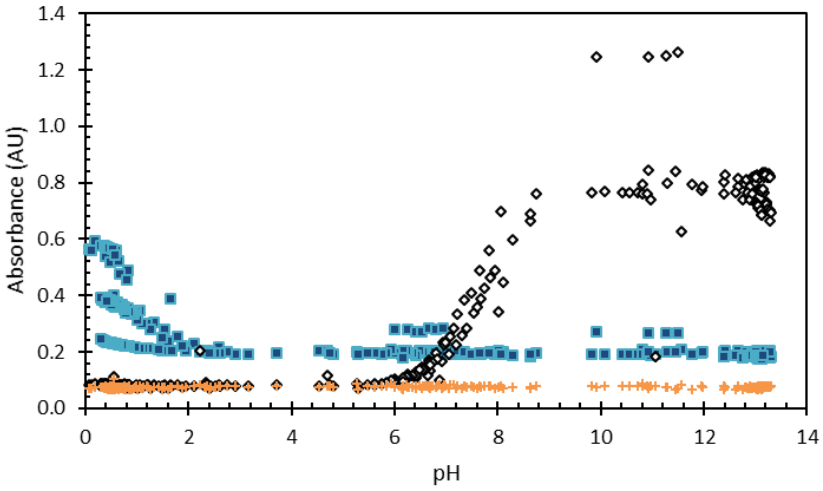


Figure 61: Absorbance Spectra of Phenol Red from pH 5.5 to 10.4. Absorbance spectra of 100 mM sodium phosphate buffer containing 20 µM phenol red at various stages of titration with hydrochloric acid. Spectra revealed an absorption maximum at 432 nm under acidic conditions and at 557 nm under alkaline conditions. Measurements were performed at room temperature (20 °C). An isosbestic point was detected at 479 nm.

A



B

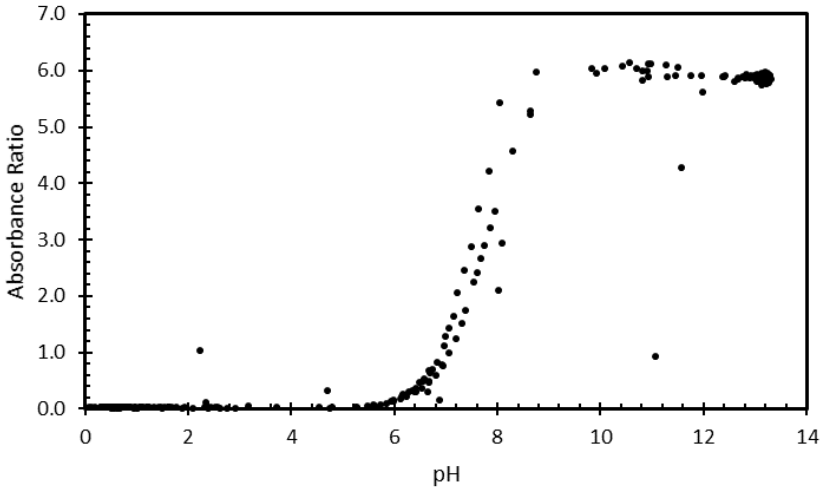


Figure 62: Absorbance vs. pH data for all of the non-enzymatic simulated conversion samples used in this study. Samples set-up parameters ranged from: 100.20 mM to 71.17 mM PO_4, with 0.00 mM to 1100.00 mM Na^+, and 0.00 mM to 1500.00 mM Cl^-, while maintaining a constant 20 µM phenol red. All samples were prepared and measured at room temperature (20 °C). (A) samples were monitored at 479 nm (blue filled squares) and 557 nm (open diamonds), and 700 nm (orange plus-signs), corresponding to measurements that vary with pH, an isosbestic point, respectively, and a high-wavelength baseline. (B) Absorbance data are normalized by subtracting the value at 700 nm from each 479 and 557 nm value, and then dividing the baseline-correct values at 557 nm by those at 479 nm. This transformation reveals a sigmoidal relationship between absorbance ratio and pH value.

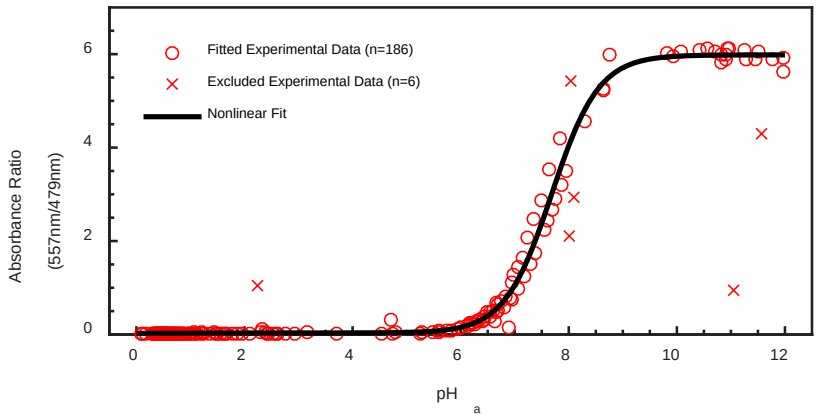


Figure 63: Nonlinear fitting of the 557 nm/479 nm absorbance ratio for phenol red. As phenol red has three protonation states, a three-state equilibrium model was fitted to the absorbance ratio data in MATLAB. Points with residuals greater than two standard deviations from the mean were labeled as outliers and excluded from a subsequent fitting routine. This process revealed the characteristic absorbance ratios of the doubly protonated, singly protonated, and fully deprotonated states to be 0.02 ± 0.03 (negligible), 0.03 ± 0.02 (negligible), and 5.98 ± 0.03, respectively (value ± 1 standard deviation).

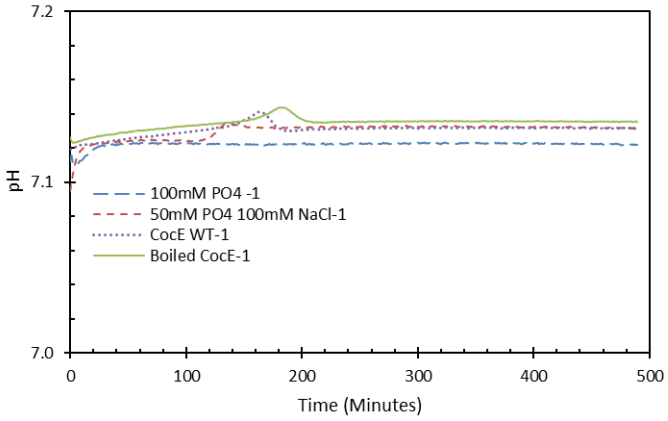


Figure 64: CocE has no activity on phenol red. Time-series data of pH, as determined by spectrophotometric monitoring of phenol red in the presence or absence of active or boiled CocE. The enzyme was shown not to have activity towards phenol red. Therefore, any measured change in pH is due to the reactivity of an active enzyme towards a substrate whose consumption results in the formation of an acid or a base, which allows for the use of phenol red as the indicator in the pH-based colorimetric assay.

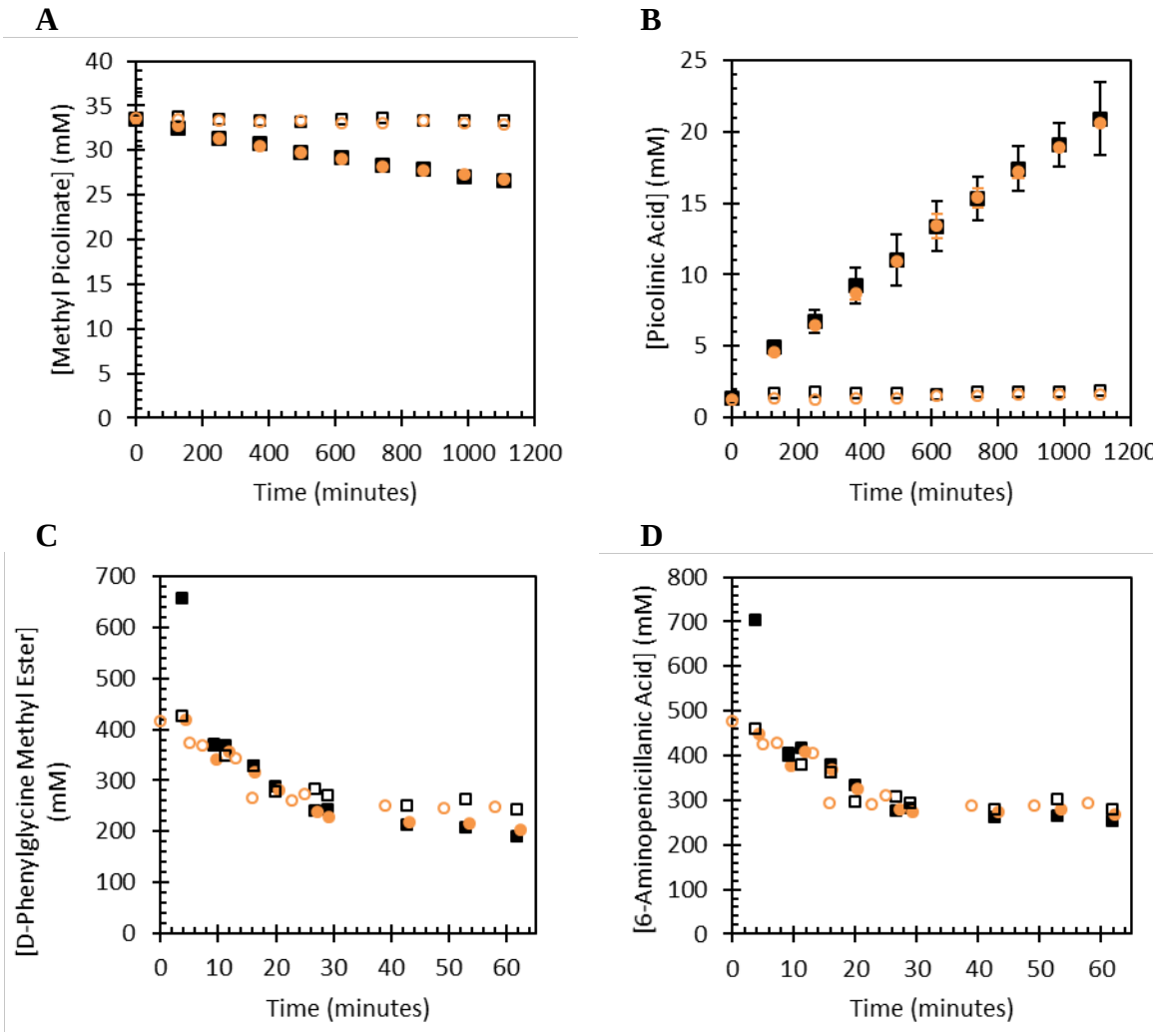


Figure 65: Phenol red is not an inhibitor of either CocE or AEH quadruple variant H (QV-H). The progress of methyl picolinate hydrolysis (A, B) and ampicillin synthesis (C-F) by CocE and AEH QV-H, respectively, were tracked by HPLC in the presence (black squares) and absence (orange circles) of phenol red. The experiments are tested at room temperature (20 °C), 100 mM PO_4 with a final reaction volume of 200 µL. Enzymatic data are shown with filled markers, and non-enzymatic controls are shown with unfilled markers. Because phenol red is not an inhibitor of CocE or the AEH variants, it can be used as the indicator for the pH-based assay. (Continued on next page)

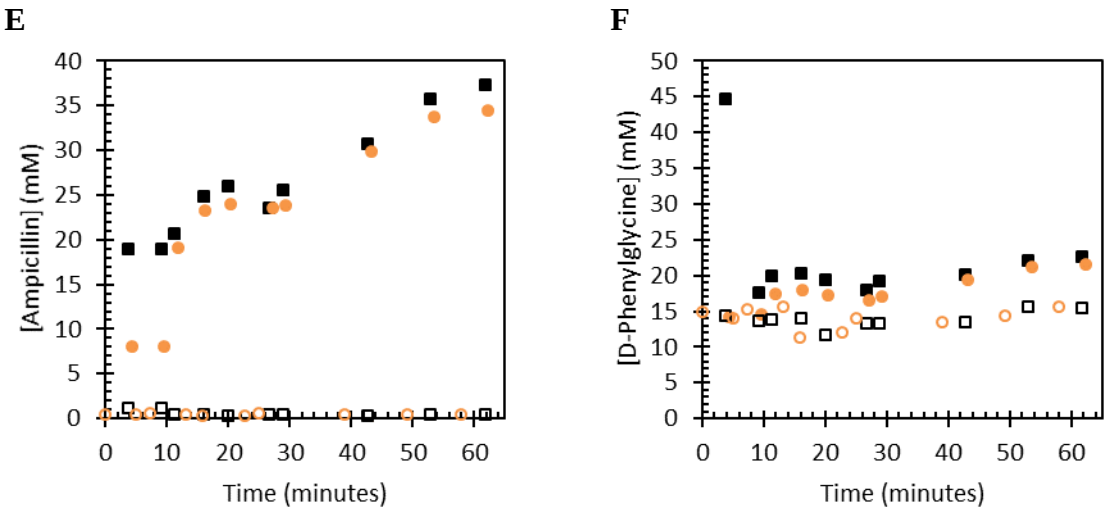


Figure 65 continued.

C.1.5 Enzyme Kinetics

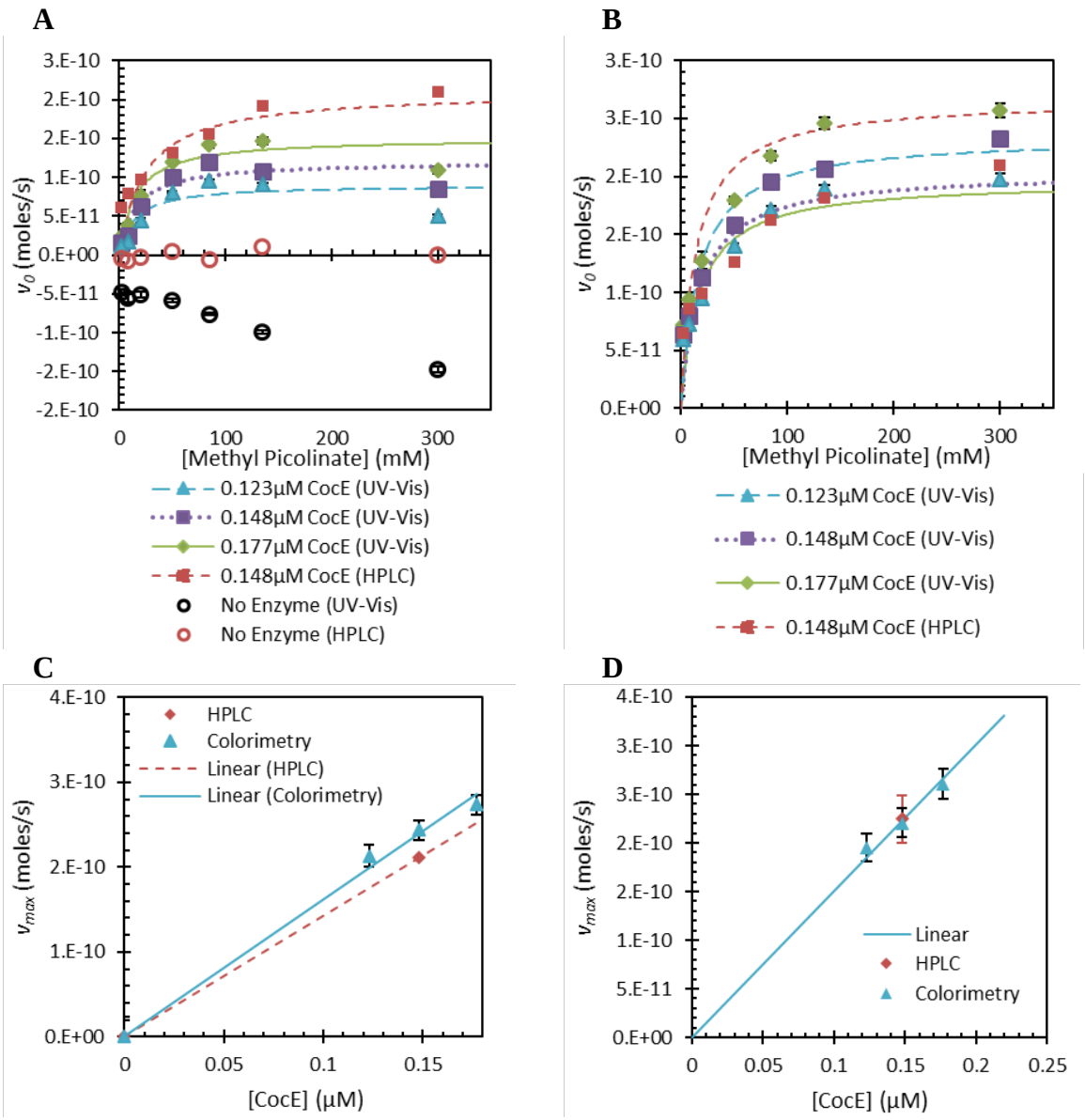


Figure 66: Michaelis-Menten kinetics for CocE WT and non-enzymatic chemical hydrolysis of methyl picolinate. The reaction setup also includes testing at room temperature (20 °C), 100 mM PO_4 with a final reaction volume of 200 µL. (A) Gross kinetics. (B) Net Kinetics (non-enzymatic values subtracted). (C) Gross V_{max} increases with increasing enzyme concentration. (D) Net V_{max} (non-enzymatic values subtracted) increases linearly with enzyme concentration. The results establish comparability between results obtained by HPLC and colorimetric methods of measurement.

Table 25: Gross kinetic parameters for the hydrolysis of methyl picolinate by CocE as determined by HPLC and spectrophotometry. Both HPLC and pH-based spectrophotometric assay kinetic data within error limit, which indicates that the latter method of study can be use in the place of HPLC. The reaction setup is room temperature (20 °C), 100 mM PO_4 with a final reaction volume of 200 µL. Initial substrate concentrations to obtain kinetics parameters are plotted in Figure 66.

Instrumentation	No.	[CocE] (µM)	K_M (mM)	V_{max} ($nmol\ s^{-1}$)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
HPLC	1	0.148	18 ± 8	0.21 ± 0.02	6.97±.07	380 ± 40
Spectrophotometer	2	0.123	14 ± 12	0.09 ± 0.02	3.5 ± 0.7	250 ± 50
Spectrophotometer	3	0.148	15 ± 8	0.12 ± 0.02	3.9 ± 0.5	260 ± 30
Spectrophotometer	4	0.177	15 ± 7	0.15 ± 0.02	4.1 ± 0.5	270 ± 30

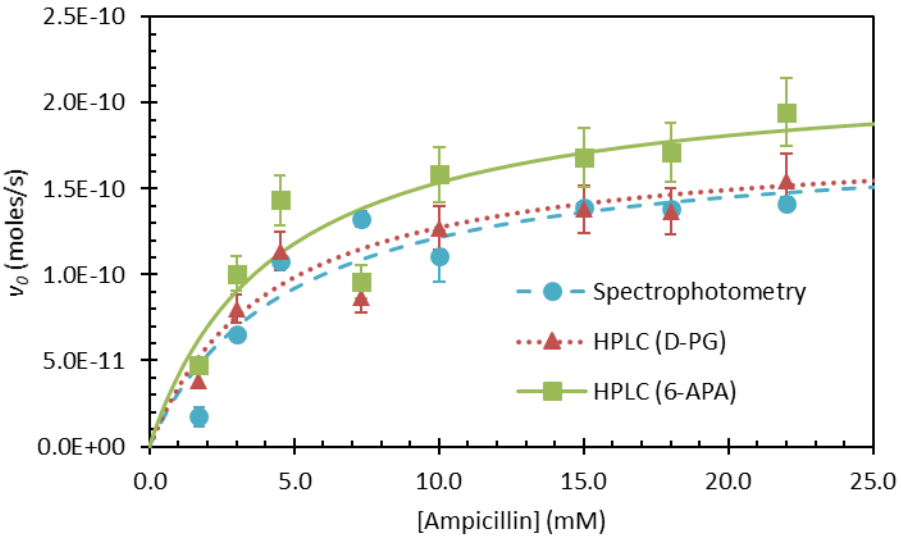


Figure 67: Michaelis-Menten kinetics for the hydrolysis of ampicillin by AEH QV-G. Kinetics determined by HPLC were obtained by independently monitoring the formation of both products, D-phenylglycine (D-PG) and 6-aminopenicillanic acid (6-APA). Including 100 mM PO_4 as the buffer at room temperature (20 °C), the reaction setup consists of a final enzyme concentration of 0.110 µM in a final reaction volume of 200 µL. The results show that both methods have comparable kinetics data, allowing for the pH-based assay to be used in the place of HPLC.

Table 26: Net kinetic parameters for the hydrolysis of ampicillin by AEH QV-G. HPLC (6-APA) is calculated using one of enzymatic reaction products: 6-APA while HPLC (D-PG) tracks the other product: D-PG. Spectrophotometry refers to measurements taken using the pH-based assay, which is dependent on the products of the enzymatic reaction. The results show that HPLC and spectrophotometric kinetics parameters are comparable and, therefore, pH-based spectrophotometric assay can be used in the place of HPLC.

Instrumentation	No.	[AEH] (μM)	K_M (mM)	V_{max} ($\mathbf{nmol\ s^{-1}}$)	k_{cat} ($\mathbf{s^{-1}}$)	k_{cat}/K_M ($\mathbf{M^{-1}s^{-1}}$)
HPLC (6-APA)	1	0.110	4±2	0.22±0.03	10±1	2000±1000
HPLC (D-PG)	2	0.110	4±1	0.18±0.02	7.9±0.9	1900±700
Spectrophotometer	3	0.110	5±2	0.18±0.02	8±1	1700±800

C.1.6 LC-MS Evidence for Peptide Synthesis

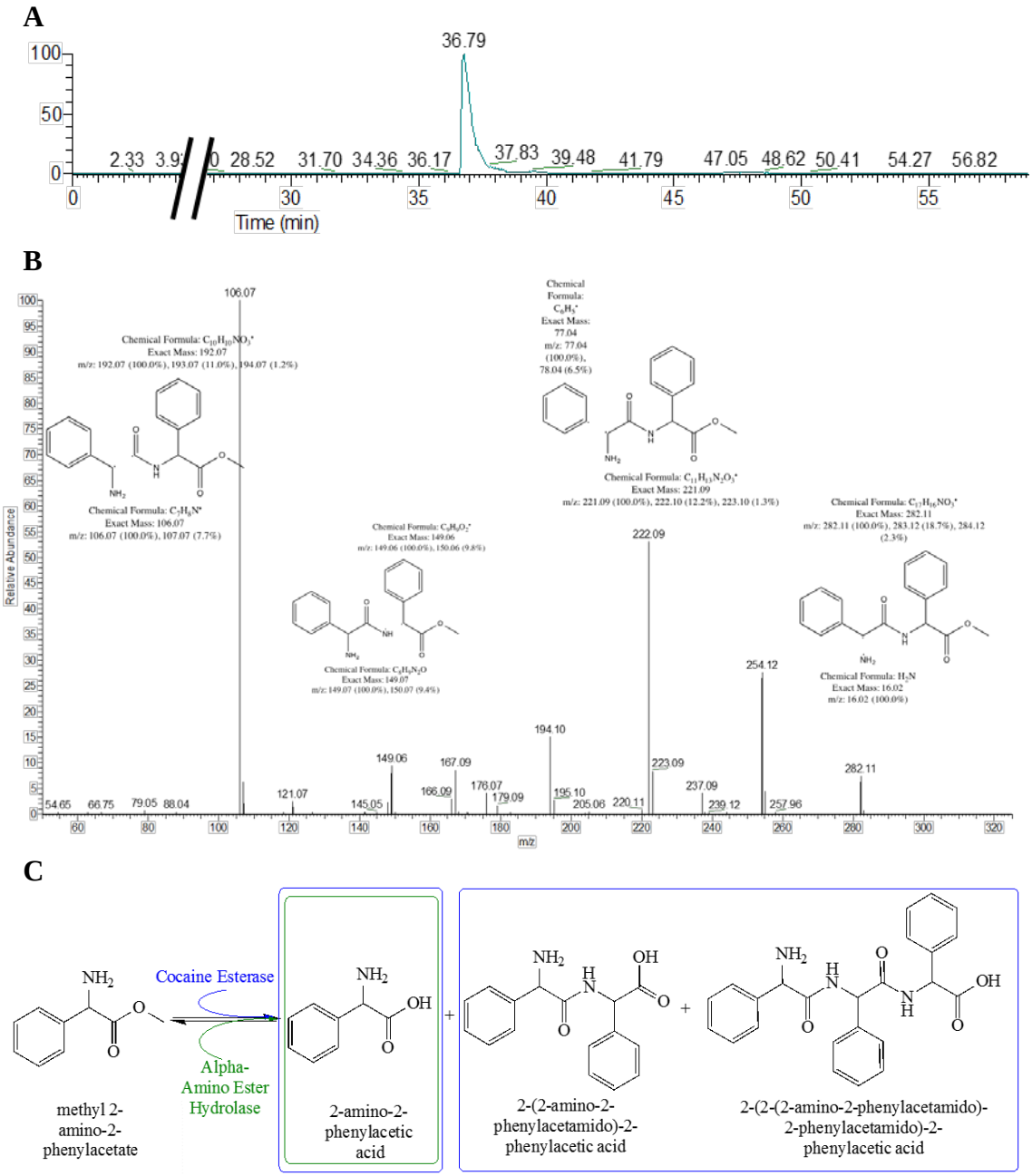


Figure 68: Tracking peptide formation over time with LC-MS. (A) HPLC chromatographs of CoCE product. (B) Fragmentation mass spectrum of species eluted from HPLC around 37 min corresponding to dipeptide formation. (C) AEH produces amino acids, while CocE produces a mixture of amino acid, as well as di- and tripeptides. All experiments were performed in triplicate.

C.2 Model Documentation

C.2.1 Modeling Quick-Start Guide

This section will help the interested researcher to use and modify the pH-based assay model.

1. **Introduction:**

   a. This model should work ‘out of the box’ on any Windows PC with Microsoft Excel and MATLAB version R2016a or later and sufficient RAM for the level of calculations being performed. Mathematica is not necessary unless the user wishes to modify the system of equations used in the model.

2. **Model Files:**

   a. `Model_File_1.nb`:

      i. This is the Mathematica notebook used to algebraically solve the system of equations described in C.2.4, Explicit System of Model Equations. More on the Mathematica Notebook itself is described in C.2.5, Mathematica Input: Solving the Model Equations and C.2.6, Mathematica Output: Expressions for Absorbance, Conversion, and Ionic Strength.

   b. `Model_File_2.m`:

i. This is the MATLAB script that runs the model. More information on the MATLAB code is provided in C.2.7, MATLAB Script Explanation. Make sure that this file is located within the MATLAB working directory.

c. `Model_File_3.xlsx`:

i. This is the spreadsheet that is used to provide the reaction and assay configuration settings to MATLAB. It also contains a sheet for curating a library of compounds and their pK_a values, as well as a sheet for providing data that should be plotted within MATLAB, sheets for receiving calculated values from MATLAB, and a sheet for hypothesis testing based on the model output and experimental data. More information on the spreadsheet is provided in C.2.3, Excel Spreadsheet.

ii. Make sure that this file is located within the MATLAB working directory. Changes in the filename of the spreadsheet must be reflected in the MATLAB file.

3. Setting up an Assay Calculation:

a. Place `Model_File_2.m` and `Model_File_3.xlsx` in the MATLAB working directory.

b. Open `Model_File_3.xlsx` and navigate to the sheet titled '`pKa Curation`'

i. In a new row, enter the name of a reactant to model. For each protonation state, enter the charge of that state.

ii. For each deprotonation step, enter the corresponding pK_a value. Using the cell drop-down menu, select the reporting basis for that event (i.e., concentration, activity, or mixed-mode), as well as the ionic strength at which the pK_a value was measured. The spreadsheet will automatically convert the pK_a values to a thermodynamic basis and calculate a value for 'z,' the charge of the most-deprotonated state, assuming that two sequential deprotonation steps are possible. 'z' is explained in more detail below.

c. Navigate to the sheet titled 'Conditions'

i. Each column in this sheet is an assay that the MATLAB script will attempt to run. There is not a fundamental limit to the number of assays that can be run from a single sheet, however subsequent calculations within the spreadsheet will stop at 10. Additional assays can always be set up and run from additional copies of the spreadsheet.

ii. Enter an identification number (e.g., 1,2,3…) into the '`AssayNum`' field.

iii. Enter a descriptive identifying title into the '`AssayTitle`' field.

iv. The model is set up to accept dilute aqueous reactions of the form: $A(+water) \rightleftharpoons B + C$, so the row-sections titled '`A`,' '`B`,' and '`C`' refer to the substrate and two products of a hydrolysis reaction. In each section, click on the cell corresponding to either '`A`,' '`B`,' or '`C`' and use the dropdown menu to select

the desired compound. The dropdown menu list is automatically populated from the list of species on the '`pKa Curation`' sheet.

v. Enter the initial concentrations of A, B, and C into the '`Ainit`,' '`Binit`,' or '`Cinit`' fields.

vi. Enter the standard deviation relating to the uncertainty of the initial concentrations of A, B, and C into the '`Ainit_std`,' '`Binit_std`,' and '`Cinit_std`' fields.

vii. Selection of a compound using the drop-down menu will auto-fill in the pK_a fields and charge number, '`z`,' field. If a pK_a does not exist, a nonphysical value of 20 is used. These values can be overridden if desired. Thermodynamic values of pK_a must be used.

1. A, B, and C can have either zero, one, or two pK_a^{th} values.

2. The charge number, '`z`' for A, B, and C are entered in their most deprotonated states, *assuming* there are two pK_a^{th} values. The following are some examples of possible situations:

   a. Species XX is neutral in its protonated state and has only one pK_a^{th}.

      i. Set pKaXX1 equal to the established pK_a^{th}value for XX.

      ii. Set pKaXX2 equal to 20, so that the calculated amount will always be nearly zero.

iii. Set zXX equal to: $(0) - 2 = \boxed{-2}$.

b. Species XX is +1 charged in its protonated state and has only 1 pK_a^{th}.

i. Set pKaXX1 equal to the established pK_a^{th} value for XX.

ii. Set pKaXX2 equal to 20, so that the calculated amount will always be nearly zero.

iii. Set zXX equal to $(+1) - 2 = \boxed{-1}$

c. Species XX is a zwitterion with two pK_as.

i. Set pKaXX1 equal to the established value for the first (lower value) pK_a^{th} of XX.

ii. Set pKaXX2 equal to the established value for the second (higher value) pK_a^{th} of XX.

iii. Set zXX equal to $(0) - 2 = \boxed{-2}$

3. Enter the standard deviation relating to the uncertainty of the pK_a of A, B, and C into each of the corresponding fields.

viii. The section beginning with the row titled '`IndicatorTot`' contains information on the pH indicator. The default indicator is phenol red. Fill in all values in this section as needed.

1. The first-time user should try running the assay with phenol red. The advanced user may adapt the model to other indicators by modifying the system of equations and rebuilding the model, or by simply ignoring the optical absorbance component of the model.

2. The user should determine which wavelength or wavelength ratio provides the best absorption signal on their own equipment over the pH range of interest, and enter those molar extinction coefficients or absorption ratios into the fields corresponding to the appropriate protonation state.

3. It is recommended to use the absorption at a wavelength which is sensitive to pH, and normalize it by dividing by the absorption at a wavelength which is insensitive to pH (the isosbestic point).

4. '`AbsIndH2`' refers to the absorbance (or absorbance ratio) of the diprotic state, '`AbsIndH`' refers to the monoprotic state, and '`AbsInd`' to the fully deprotonated state. '`AbsIndH2_Std`' etc. refer to the standard deviations corresponding to the uncertainty of the absorbance values.

ix. The section beginning with the row titled '`Phostot`' contains information on the buffer. Fill in all values in this section as needed.

    1. The default buffer is phosphate buffer, which is triprotic.

    2. The default values of pK_a^{th} for phosphate buffer should be sufficient for most room temperature applications.

3. Other buffers can be used by modifying the pK_a^{th} values for phosphate, keeping in mind that if the charge numbers differ from those of phosphate buffer, the system of model equations should be re-solved and the model adjusted accordingly.

x. The section beginning with the row titled '`pKw`' contains the thermodynamic value of the autoionization constant of water, and its uncertainty.

xi. The sections beginning with rows titled '`Na`' and 'Cl' contain information on the concentrations of Na^+ and Cl^- ions respectively, assuming complete dissociation. These values must be as close to the true experimental values as possible. Mathematically speaking, an error in these values is like an error in addition of NaOH or HCl to the assay vessel, and will affect pH calculations accordingly. If it is suspected that there are errors in the experimental values of $[Na^+]$ or $[Cl^-]$, iteratively adjust the value of one or the other and run the simulation until the predictions match the experimentally observed pH.

xii. If *simulated* (i.e., non-enzymatic data simulating a range of conversion values) exist, enter the data for Absorbance, Conversion, and pH measurements into the sheets labeled '`I_A`,' '`I_X`,' and '`I_pH`,' ('`I_...`' for *input*) respectively, in the column corresponding to the set of conditions that were simulated. These data will be imported into MATLAB and plotted along with the simulated curves.

xiii. If experimental time-series assay data exist (i.e., enzymatic data and non-enzymatic controls), enter them in the appropriate columns of the sheet labeled ‘Hypothesis Testing.’ If the data are not collected at the same timepoints, offset one column to match the timepoints for the enzymatic data and controls as closely as possible.

d. Save and close the spreadsheet.

e. Open the MATLAB script file.

i. Set the value of `MC_samps` equal to the desired number of Monte Carlo samples. Higher values are more computationally expensive and time consuming. For quick setup and troubleshooting, a value of zero is preferred. For meaningful statistical analysis and hypothesis testing, a value of 10000 gives good results.

ii. Set the value of `MoreGraphs` equal to ‘true’ or ‘false (no quotation marks) to toggle the generation and display of additional plots and information.

iii. Set the value of `WriteToExcel` equal to ‘true’ or ‘false (no quotation marks) to toggle sending output to Excel. If this is set to ‘true,’ tabulated predictions for the relationship between the absorbance and conversion of each set of assay conditions will be automatically saved to the Excel Spreadsheet in the sheets titled ‘`O_A(pH)_Mean,`’ ‘`O_X(A)_Mean`,’ and ‘`O_X(A)_Std`’ (‘`O_`...’ for *output*). This is necessary for hypothesis testing.

1. NOTE: It is always advisable to save an un-run copy of the spreadsheet in case the model needs to be re-run with different parameters. The MATLAB script does not clear the contents of ‘`O_A(pH)_Mean`,’ ‘`O_X(A)_Mean`,’ and ‘`O_X(A)_Std`,’ so having a fresh copy handy eliminates the need to manually erase values from prior code executions.

iv. Set the values of `pH_samps`, `pH_LowerLimit`, and `pH_UpperLimit`. For example, a `pH_samps` value of 1000 should provide sufficient resolution between lower and upper limits of 4 and 8, respectively.

v. Set the value of `ExcelFile` to reflect the filename of the Excel Spreadsheet containing the assay parameters and data (e.g., `'Model_File_3.xlsx'`; *with single quotes*).

vi. Save the MATLAB script and close any other open files or documents on the computer.

4. Executing a MATLAB calculation:

a. Run the MATLAB script by pressing the ‘Run’ button, or by pressing ‘F5’ on the keyboard when the script editor window is open.

i. To abort the calculation at any time, mouseclick in the MATLAB command window and press ‘Ctrl’ + ‘c’ on the keyboard.

b. Data may be output in three locations:

i. The Command Window will show updates on the progress of the calculations.

ii. If `MoreGraphs` was set to 'true,' new figures with plots will pop up periodically as the calculation progresses.

iii. If `WriteToExcel` was set to 'true,' the calculations will be saved to the referenced Excel file.

5. Helpful Tips:

a. Note that small errors in the concentration of salt can have big effects on the model. Due to the charge balance, a bit of extra Na^+ without the matching Cl^- implies the addition of NaOH. Similarly, a bit of extra Cl^- without the matching Na^+ implies the addition of HCl.

b. If the model throws an error and Monte Carlo sampling is not being performed:

i. Verify that it is possible for the specified input parameters to produce values in the pH range that is being requested, or that the pH resolution is not set too coarsely. Many input pH values may be trimmed significantly, in order to find the range where conversion takes physically meaningful values.

ii. Verify that the reaction is expected to produce a deviation in pH. If no pH change occurs, then the starting pH could correspond to any value of conversion, and the assay response is mathematically undefined.

c. If the model throws an error and Monte Carlo sampling is being performed, ensure that enough samples are being used, and that it is reasonable to suspect that the input parameter(s) being varied would alter the outcome of the calculations. For example, a small error on a pK_a^* near 11 cannot be expected to produce significant error if the model is only being operated between pH 4 and 7. Too small an error will cause problems for the standard deviation fitting routine because the distribution of error will be undefined.

C.2.2 Standardization of pK_a Values from the Literature

C.2.2.1 Mathematical Considerations

For a weak acid with *j* labile protons and a conjugate base of charge z, the following reversible association reaction may be written: $H^+ + AH_{j-1}^z \rightleftharpoons AH_j^{z+1}$. In general, the equilibrium positions of such reactions are reported as association or stability constants: K_a, dissociation constants: $K_d = K_a^{-1}$, or pK_a values: $pK_a = \log_{10}(K_a) = -\log_{10}(K_d)$. When ionic strength effects are neglected, stability constants are written on a concentration basis, denoted K_a^c and pK_a^c, where the superscript 'c' is used in deference to the use of concentration:

$$K_a^c = \frac{\left[AH_j^{z+1}\right]}{[H^+]\left[AH_{j-1}^z\right]} \rightarrow pK_a^c = \log_{10}(K_a^c) \tag{68}$$

However, unless calibrated using special procedures as discussed in CHAPTER 5, section 5.3.4.3, Potentiometric Electrode Calibration, typical pH electrodes report proton *activity*, and not proton *concentration*. This leads to the use of the "mixed-mode" stability constant, also referred to as "apparent," or "practical" and denoted K_a^*, according to the terminology of Ellis and Morrison (1982) [186]:

$$K_a^* = \frac{\left[AH_j^{z+1}\right]}{(a_{H^+})\left[AH_{j-1}^z\right]} \rightarrow pK_a^* = \log_{10}(K_a^*) \tag{69}$$

where a_{H^+} is the thermodynamic activity of H^+ in solution.

Yet another way to report these values (and one which is independent of ionic strength) is on an activity basis. Such values are referred to as the "true" or "thermodynamic" constant, and are indicated as such by a superscript "th":

$$K_a^{th} = \frac{\left(a_{AH_j^{z+1}}\right)}{(a_{H^+})\left(a_{AH_{j-1}^{z}}\right)} \rightarrow pK_a^{th} = \log_{10}(K_a^{th}) \tag{70}$$

The thermodynamic activity of species i, a_i, is the product of the species concentration c_i and the activity coefficient γ_i. Therefore, equilibrium constants reported on a concentration basis are related to mixed-mode or thermodynamic values by a ratio of activity coefficients $K_a^{th} = K_a^c \times K_a^{\gamma c}$, or $K_a^{th} = K_a^* \times K_a^{\gamma *}$, where

$$K_a^{\gamma c} = \frac{\left(\gamma_{AH_j^{z+1}}\right)}{(\gamma_{H^+})\left(\gamma_{AH_{j-1}^{z}}\right)}, \qquad K_a^{\gamma *} = \frac{\left(\gamma_{AH_j^{z+1}}\right)}{\left(\gamma_{AH_{j-1}^{z}}\right)} \tag{71}$$

In general, the value of γ_i represents the extent to which thermodynamic activity of i deviates from ideality, and is a function of the charge of i, and the ionic strength, I, of the bulk solution. For the purposes of this work, it is essential to standardize acid-base equilibrium constants pulled from the literature into a useful collection with a consistent basis so that ionic strength corrections can be applied consistently to all species. To that end, the extended Debye Hückel relationship presented by Sun, et al. (1980) [156] is used to extrapolate mixed-mode and concentration-based equilibrium constants to infinite dilution.

Here, we present an adaptation of their derivation. The ionic activity coefficient in of a single ion of charge ζ in aqueous solution can be estimated with the following relationship:

$$\log_{10}(\gamma_i) = -0.5\zeta^2\left(\frac{\sqrt{I}}{1+\sqrt{I}}\right) + 0.1\zeta^2 I = \zeta^2\left(-0.5\left(\frac{\sqrt{I}}{1+\sqrt{I}}\right) + 0.1I\right) \quad (72)$$

Applying this relationship to the equilibrium reaction $r_1 + r_2 \rightleftharpoons p_1$ yields the following lumped activity coefficient term:

$$pK_a^{\gamma} = \log_{10}\left(\frac{(\gamma_{p1})}{(\gamma_{r1})(\gamma_{r2})}\right) = \log_{10}(\gamma_{p1}) - \log_{10}(\gamma_{r1}) - \log_{10}(\gamma_{r2}) \quad (73)$$

By combining like terms, this can be simplified to the following form:

$$pK_a^{\gamma} = \phi\left(-0.5\left(\frac{\sqrt{I}}{1+\sqrt{I}}\right) + 0.1I\right) \quad (74)$$

where $\phi = \left(\zeta_{p1}^2 - \zeta_{r1}^2 - \zeta_{r2}^2\right)$. In terms used by Sun et al, this relationship is written as:

$$\Delta \log K = pK_a^{\gamma} = (\chi)(0.51)\left(\frac{\sqrt{I}}{1+A\sqrt{I}} + BI\right) + CI \quad (75)$$

$$\chi = -\phi = \left(\zeta_{r1}^2 + \zeta_{r2}^2 - \zeta_{p1}^2\right) \quad (76)$$

where A, B, and C are empirically fitted constant parameters equal to 1.50, -0.09, and -0.09, respectively.

For pK_a values reported on a concentration basis, this lumped correction becomes:

$$\chi = z^2 + 1 - (z+1)^2 = -2z \tag{77}$$

$$pK_a^{\gamma c} = (-2z)(0.51)\left(\frac{\sqrt{I}}{1 + A\sqrt{I}} + BI\right) + CI \tag{78}$$

while the correction for pK_a values reported on a mixed-mode basis becomes:

$$\chi = z^2 - (z+1)^2 = -2z - 1 \tag{79}$$

$$pK_a^{\gamma *} = (-2z - 1)(0.51)\left(\frac{\sqrt{I}}{1 + A\sqrt{I}} + BI\right) + CI \tag{80}$$

By putting these pieces together, we can now collect pK_a values from the literature in various forms and convert them to thermodynamic values with the following:

$$pK_a^{th} = \log_{10}\left(K_a^c \times K_a^{\gamma c}\right) = pK_a^c + pK_a^{\gamma c} \tag{81}$$

$$pK_a^{th} = \log_{10}\left(K_a^* \times K_a^{\gamma *}\right) = pK_a^* + pK_a^{\gamma *} \tag{82}$$

Where Equation (81) is used if the literature data are reported on a concentration basis, and Equation (82) is used if the literature data are reported on a mixed-mode basis.

These thermodynamic values are then supplied to MATLAB and iteratively corrected to a concentration-basis at the appropriate ionic strength using the relationship in Equation (83).

$$pK_a^c = pk_a^{th} - pK_a^{\gamma c} \tag{83}$$

C.2.2.2 Charges, Protonation States, and Curated pK_a Values for Model Species

Table 27: Curated pK_a values and their references. This table contains the pK_a values that were used in modeling, as well as the reference for each value, the ionic strength reportedly used for measurement, and the apparent basis of the value (i.e., concentration, activity, or mixed-mode). The table also contains the extrapolated thermodynamic values, pK_a^{th} which were calculated as described previously.

	First Deprotonation Event			Second Deprotonation Event			Third Deprotonation Event			Extrapolated to Thermodynamic Basis		
Species	$pK_{a,1}$	I_m	Ref	$pK_{a,2}$	I_m	Ref	$pK_{a,3}$	I_m	Ref	$pK_{a,1}^{th}$	$pK_{a,1}^{th}$	$pK_{a,1}^{th}$
Phosphoric acid	1.924[C]	0.1	211	6.716[C]	0.1	211	11.523[C]	0.1	211	2.12	7.13	12.1
Phenol red	1.03[MM]	0.1	268	7.70[C]	0.1	211				1.13	8.11	
Methyl picolinate	2.21[A]	0	269-							2.21		
Picolinic acid	1.01[MM]	0.2	112	5.29[MM]	0.2	112				0.86	5.40	
Methanol	15.3[A]	0	271							15.30		
Phenylglycine methyl ester	7.24[A]	0	267							7.24		
Phenylglycine	1.90[C]	0.1	211	8.92[C]	0.1	211				1.89	9.12	
Acetamide	-0.37[A]	0	112							-0.37		
Acetic acid	4.756[A]	0	112							4.76		
Ammonium	9.24[A]	0	112							9.24		
Nicotinamide	0.50[A]	0	272	3.33[A]	0	112				0.50	3.33	
Nicotinic acid	2.07[A]	0	112	4.75[A]	0	112				2.07	4.75	
Ampicillin	2.14[A]	0	267	7.31[A]	0	267				2.14	7.31	
6-aminopenicillinic acid	2.20[A]	0	267	4.83[A]	0	267				2.20	4.83	

[C]: Concentration basis; [A]: Activity basis; [MM]: Mixed-mode.

C.2.3 Excel Spreadsheet

A sample spreadsheet, `Model_File_3.xlsx`, is also provided. The spreadsheet consists of 9 sheets:

1. `pKa Curation`**:** Organizes and standardizes literature values for pK_as of different compounds.
2. `Conditions`**:** Inputs assay conditions that will be read in by the MATLAB model.
3. `I_A`**:** Inputs absorbance ratio data of simulated reaction controls, for overlay with predictions in MATLAB.
4. `I_X`**:** Inputs conversion data of simulated reaction controls, for overlay with predictions in MATLAB.
5. `I_pH`**:** Inputs pH_a data of simulated reaction controls, for overlay with predictions in MATLAB.
6. `O_A(pH)_Mean`**:** Receives MATLAB predictions of nominal or mean absorbance as a function of measured assay pH_a.
7. `O_X(A)_Mean`**:** Receives MATLAB predictions of nominal or mean conversion as a function of measured absorbance.
8. `O_X(A)_Std`**:** Receives MATLAB predictions of the standard deviation on mean conversion as function of absorbance.
9. `Hypothesis Testing`**:** Performs Welch's t-test to compare pairs of experimental absorbance vs. time data.

C.2.4 Explicit System of Model Equations

This section explains the full system of equations used to represent hydrolysis, solution pH, and indicator absorbance.

Equations (84)-(86) describe the ionization of the reactant, A. Equation (84) describes the first ionization of A. If not applicable, set $pK_{aA1} = 20$. Equation (85) describes the second ionization of A. If not applicable, set $pK_{aA2} = 20$. Equation (86) describes the analytical (total) concentration of the conjugate base of A in solution.

$$[AH_2] \times 10^{-pK_{aA1}} = [AH] \times 10^{-pH} \tag{84}$$

$$[AH] \times 10^{-pK_{aA2}} = [A] \times 10^{-pH} \tag{85}$$

$$[A]_{tot} = [AH_2] + [AH] + [A] \tag{86}$$

Equations (87)-(89) describe the ionization of the first hydrolysis product, B. Equation (87) describes the first ionization of B. If not applicable, set $pK_{aB1} = 20$. Equation (88) describes the second ionization of B. If not applicable, set $pK_{aB2} = 20$. Equation (89) describes the analytical (total) concentration of the conjugate base of B in solution.

$$[BH_2] \times 10^{-pK_{aB1}} = [BH] \times 10^{-pH} \tag{87}$$

$$[BH] \times 10^{-pK_{aB2}} = [B] \times 10^{-pH} \tag{88}$$

$$[B]_{tot} = [BH_2] + [BH] + [B] \tag{89}$$

Equations (90)-(92) describe the ionization of the second hydrolysis product, C. Equation (90) describes the first ionization of C. If not applicable, set $pK_{aC1} = 20$. Equation (91) describes the second ionization of C. If not applicable, set $pK_{aC2} = 20$. Equation (92) describes the analytical (total) concentration of the conjugate base of C in solution.

$$[CH_2] \times 10^{-pK_{aC1}} = [CH] \times 10^{-pH} \tag{90}$$

$$[CH] \times 10^{-pK_{aC2}} = [C] \times 10^{-pH} \tag{91}$$

$$[B]_{tot} = [BH_2] + [BH] + [B] \tag{92}$$

Equations (93)-(96) describe the ionization of the phosphate buffer. Equations (93), (94), and (95) describe the first, second, and third ionization steps of phosphoric acid. Equation (96) describes the total concentration of phosphate in solution.

$$[H_3PO_4] \times 10^{-pK_{aPhos1}} = [H_2PO_4] \times 10^{-pH} \tag{93}$$

$$[H_2PO_4] \times 10^{-pK_{aPhos2}} = [HPO_4] \times 10^{-pH} \tag{94}$$

$$[HPO_4] \times 10^{-pK_{aPhos3}} = [PO_4] \times 10^{-pH} \tag{95}$$

$$[Phos]_{tot} = [H_3PO_4] + [H_2PO_4] + [HPO_4] + [PO_4] \tag{96}$$

Equations (97)-(100) describe the absorbance (at a particular wavelength, or a ratio of wavelengths) of a halochromic indicator. Equations (97) and (98) describe the first and second deprotonation steps of the indicator, respectively. Equation (99) describes the total absorbance of indicator in solution. Equation (100) describes the overall absorbance of the sample as a linear combination of the absorbances of each individual protonation state. Phenol red was used as the indicator in this work, and absorbance refers to the ratio of measured absorbance at 557 nm/479 nm.

$$[IndH_2] \times 10^{-pK_{aInd1}} = [IndH] \times 10^{-pH} \tag{97}$$

$$[IndH] \times 10^{-pK_{aInd2}} = [Ind] \times 10^{-pH} \tag{98}$$

$$[Ind]_{tot} = [IndH_2] + [IndH] + [Ind] \tag{99}$$

$$Absorbance = \frac{AbsIndH_2 \times [IndH_2] + AbsIndH \times [IndH] + AbsInd \times [Ind]}{[Ind]_{tot}} \tag{100}$$

Equation (101) describes the self-ionization of water with ionization constant pK_w.

$$10^{-pK_w} = [OH] \times 10^{-pH} \tag{101}$$

Equation (102) is a charge balance, and ensures that the total charge due to positive ions in solution exactly equals the total charge due to negative ions in solution. Z_A, Z_B, and Z_C are the charges of the least-protonated (or, most basic) states of substrate A, and

products B and C, respectively. These must be set *as if* each species has two protonation states. E.g., if A is neutral when protonated and has only one pK_a, use $Z_A = -2$ anyway.

$$\begin{aligned} 0 = 10^{-pH} + [Na] - [OH] - [Cl] - [H_2PO_4] - 2 \times [HPO_4] \\ - 3 \times [PO_4] - [IndH] - 2 \times [Ind] + (z_A + 2) \times [AH_2] \\ + (z_A + 1) \times [AH] + z_A \times [A] + (z_B + 2) \times [BH_2] \\ + (z_B + 1) \times [BH] + z_B \times [B] + (z_C + 2) \times [CH_2] \\ + (z_C + 1) \times [CH] + z_C \times [C] \end{aligned} \tag{102}$$

Equation (103) defines the ionic strength of a sample, which is $1/2$ the sum of the concentrations of all species times the square of their charges.

$$\begin{aligned} IS = \frac{1}{2}(10^{-pH} + [Na] + [OH] + [Cl] + [H_2PO_4] + 4 \times [HPO_4] \\ + 9 \times [PO_4] + [IndH] + 4 \times [Ind] \\ + (z_A + 2)^2 \times [AH_2] + (z_A + 1)^2 \times [AH] + z_A^2 \times [A] \\ + (z_B + 2)^2 \times [BH_2] + (z_B + 1)^2 \times [BH] + z_B^2 \times [B] \\ + (z_C + 2)^2 \times [CH_2] + (z_C + 1)^2 \times [CH] + z_C^2 \times [C]) \end{aligned} \tag{103}$$

C.2.5 Mathematica Input: Solving the Model Equations

The following text may be entered (or copied and pasted) into Wolfram Mathematica to produce the analytical solution to the system of model equations:

```
(* SET Global Assumptions *)
$Assumptions={{Atot,AH2,AH,A,Btot,BH2,BH,B,Ctot,CH2,CH,Cconc,Phostot,H3
PO4,H2PO4,HPO4,PO4,IndicatorH2,IndicatorH,Indicator,IndicatorTot,Absorb
ance,AbsIndH2,AbsIndH,AbsInd,OH,Na,Cl,Ainit,Binit,Cinit,pKw,IS}>=0,{Ato
t,AH2,AH,A,Btot,BH2,BH,B,Ctot,CH2,CH,Cconc,Phostot,H3PO4,H2PO4,HPO4,PO4
,IndicatorH2,IndicatorH,Indicator,IndicatorTot,Absorbance,AbsIndH2,AbsI
ndH,AbsInd,OH,Na,Cl,Ainit,Binit,Cinit,pKw,pH,X,IS}∈Reals,0<=X<=1,{zA,zB
,zC}∈Integers}
(* THREE IONIZATION STATES for SUBSTRATE (A) *)
f1 = AH2*(10^(-pKaA1)) - AH*(10^(-pH));
f2 = AH*(10^(-pKaA2)) - A*(10^(-pH));
f3 = Atot - (AH2+AH + A);
(* THREE IONIZATION STATES for PRODUCT 1 (B) *)
f4 = BH2*(10^(-pKaB1))- BH*(10^(-pH));
f5 = BH*(10^(-pKaB2))- B*(10^(-pH));
f6 = Btot - (BH2 + BH + B);
(* THREE IONIZATION STATES for PRODUCT 2 (C) *)
f7 = CH2*(10^(-pKaC1))- CH*(10^(-pH));
f8 = CH*(10^(-pKaC2))- Cconc*(10^(-pH));
f9 = Ctot - (CH2 + CH + Cconc);
(* FOUR IONIZATION STATES for PHOSPHATE BUFFER *)
f10 = H3PO4*(10^(-pKaPhos1)) - H2PO4*(10^(-pH));
f11 = H2PO4*(10^(-pKaPhos2)) - HPO4*(10^(-pH));
f12 = HPO4*(10^(-pKaPhos3))- PO4*(10^(-pH));
f13 = Phostot - (H3PO4 + H2PO4 + HPO4 + PO4);
(* THREE IONIZATION STATES for pH INDICATOR *)
f14 = IndicatorH2*(10^(-pKaIndicator1))- IndicatorH*(10^(-pH));
f15 = IndicatorH*(10^(-pKaIndicator2))- Indicator*(10^(-pH));
f16 = IndicatorTot-(IndicatorH2+IndicatorH+Indicator);
f17 = Absorbance *IndicatorTot-
(AbsIndH2*IndicatorH2+AbsIndH*IndicatorH+AbsInd*Indicator);
(* TWO IONIZATION STATES for WATER *)
f18 = OH*(10^(-pH)) - (10^(-pKw));
(* CHARGE BALANCE *)
f19 = (10^(-pH)+Na-OH-Cl-H2PO4-2*HPO4-3*PO4-IndicatorH-
2*Indicator+(zA+2)*AH2+(zA+1)*AH+zA*A+(zB+2)*BH2+(zB+1)*BH+zB*B+(zC+2)*
CH2+(zC+1)*CH+zC*Cconc);
(* REACTION STOICHIOMETRY and CONVERSION *)
f20 = Atot - Ainit*(1 - X);
f21 = Btot - (Binit + Ainit*X);
f22 = Ctot - (Cinit + Ainit*X);
(* Calculate Ionic Strength) *)
f23 = IS-
(1/2)*((AH2*(zA+2)^2)+(AH*(zA+1)^2)+(A*(zA)^2)+(BH2*(zB+2)^2)+(BH*(zB+1
)^2)+(B*(zB)^2)+(CH2*(zC+2)^2)+(CH*(zC+1)^2)+(Cconc*(zC)^2)+(H2PO4*(-
```

```
1)^2)+(HPO4*(-2)^2)+(PO4*(-3)^2)+(IndicatorH*(-1)^2)+(Indicator*(-
2)^2)+(10^(-pH))+Na+Cl+OH);
(* SOLVE SYSTEM of EQUATIONS for Conversion *)
Qx = Solve[f1 == 0 && f2 == 0 && f3 == 0 && f4 == 0 && f5 == 0 && f6 ==
0 && f7 == 0 && f8 == 0 && f9 == 0 && f10 == 0 && f11 == 0 && f12 == 0
&& f13 == 0 && f14 == 0 &&  f15 == 0 &&  f16 == 0 && f17 == 0 && f18 ==
0 && f19 == 0 && f20 == 0 && f21 == 0 && f22 == 0  &&
f23==0,{X},{Atot,AH2,AH,A,Btot,BH2,BH,B,Ctot,CH2,CH,Cconc,H3PO4,H2PO4,H
PO4,PO4,OH,Absorbance,IndicatorH2,IndicatorH,Indicator,IS}]
(* SOLVE SYSTEM of EQUATIONS for ABSORBANCE *)
Solnabs = Solve[f1 == 0 && f2 == 0 && f3 == 0 && f4 == 0 && f5 == 0 &&
f6 == 0 && f7 == 0 && f8 == 0 && f9 == 0 && f10 == 0 && f11 == 0 && f12
== 0 && f13 == 0 && f14 == 0 &&  f15 == 0 &&  f16 == 0 && f17 == 0 &&
f18 == 0 && f19 == 0 && f20 == 0 && f21 == 0 && f22 == 0
,{Absorbance},{Atot,AH2,AH,A,Btot,BH2,BH,B,Ctot,CH2,CH,Cconc,H3PO4,H2PO
4,HPO4,PO4,OH,X,IndicatorH2,IndicatorH,Indicator,IS}]
(* SIMPLIFY the EXPRESSION for CONVERSION *)
SolnX=Simplify[Qx[[1]][[1]][[2]],TimeConstraint->Infinity]
(* SOLVE SYSTEM of EQUATIONS for IONIC STRENGTH *)
Qis = Solve[f1 == 0 && f2 == 0 && f3 == 0 && f4 == 0 && f5 == 0 && f6
== 0 && f7 == 0 && f8 == 0 && f9 == 0 && f10 == 0 && f11 == 0 && f12 ==
0 && f13 == 0 && f14 == 0 &&  f15 == 0 &&  f16 == 0 && f17 == 0 && f18
== 0 && f19 == 0 && f20 == 0 && f21 == 0 && f22 == 0 &&
f23==0,{IS},{Atot,AH2,AH,A,Btot,BH2,BH,B,Ctot,CH2,CH,Cconc,H3PO4,H2PO4,
HPO4,PO4,OH,X,Absorbance,IndicatorH2,IndicatorH,Indicator}]
(* SIMPLIFY the EXPRESSION for IONIC STRENGTH *)
SolnIS = Simplify[Qis[[1]][[1]][[2]],TimeConstraint->Infinity]
```

C.2.6 Mathematica Output: Expressions for Absorbance, Conversion, and Ionic Strength

When the system of equations is evaluated algebraically in Mathematica, the following simplified expressions for absorbance, conversion, and ionic strength are produced:

C.2.6.1 Absorbance

```
Absorbance =
(10^(2*pH)*AbsInd+10^(pH+pKaIndicator2)*AbsIndH+10^(pKaIndicator1+pKaIn
dicator2)*AbsIndH2)/(10^(2*pH)+10^(pH+pKaIndicator2)+10^(pKaIndicator1+
pKaIndicator2))
```

C.2.6.2 Conversion

```
Conversion =
(10^(pKaA2+pKaB2+pKaC2+pKaIndicator1+pKaIndicator2)*(10^(2*pH+pKaC1)*(1
0^(-2*pH)+(10^(-pH)+10^(-
pKaC2))/10^pKaC1)*(10^(2*pH+pKaA1+pKaB1)*(10^(-2*pH)+(10^(-pH)+10^(-
pKaB2))/10^pKaB1)*(1000^pH*(10^(-2*pH)+(10^(-pH)+10^(-
pKaA2))/10^pKaA1)*(10^(-pKaIndicator1-
pKaIndicator2)*(2^(1+pH)*5^pH+10^pKaIndicator2)*(10^(2*pH+pKaPhos3)+10^
(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Indicat
orTot+(10^(-2*pH)+10^(-pH-pKaIndicator1)+10^(-pKaIndicator1-
pKaIndicator2))*(((10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKa
Phos1+pKaPhos2+pKaPhos3)+1000^pH)*(-10^pKw+100^pH+10^(pH+pKw)*Cl-
10^(pH+pKw)*Na))/10^pKw+100^pH*(2^(1+pH+pKaPhos3)*5^(pH+pKaPhos3)+10^(p
KaPhos2+pKaPhos3)+3*100^pH)*Phostot))-10^(-pKaA1-pKaA2-pKaIndicator1-
pKaIndicator2)*(10^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaIndicator2)+
100^pH)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaP
hos2+pKaPhos3)+1000^pH)*Ainit*(10^pKaA2*(2^(1+pKaA1)*5^pKaA1+10^pH)+(10
^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*zA))-10^(-pKaA2-pKaB2-
pKaIndicator1-
pKaIndicator2)*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaIndic
ator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH+pKaPhos3)+10^
(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Binit*(
10^pKaB2*(2^(1+pKaB1)*5^pKaB1+10^pH)+(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+10
0^pH)*zB))-10^(-pKaA2-pKaB2-pKaC2-pKaIndicator1-
pKaIndicator2)*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+1
0^(pKaB1+pKaB2)+100^pH)*(10^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaInd
icator2)+100^pH)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaP
hos1+pKaPhos2+pKaPhos3)+1000^pH)*Cinit*(10^pKaC2*(2^(1+pKaC1)*5^pKaC1+1
0^pH)+(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zC)))/((10^(pH+pKaIndicat
```

```
or2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH+pKaPhos3)+10^(p
H+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Ainit*((-
2^(1+4*pH+pKaA1+pKaA2))*5^(4*pH+pKaA1+pKaA2)+2^(1+4*pH+pKaB1+pKaB2)*5^(
4*pH+pKaB1+pKaB2)+2^(1+4*pH+pKaB2+pKaC2)*5^(4*pH+pKaB2+pKaC2)+2^(1+2*pH
+pKaA2+pKaB1+pKaB2+pKaC2)*5^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)+2^(1+4*pH+pK
aC1+pKaC2)*5^(4*pH+pKaC1+pKaC2)+2^(1+2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*5^(2
*pH+pKaA2+pKaB2+pKaC1+pKaC2)+2^(2+2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*5^(2*pH
+pKaB1+pKaB2+pKaC1+pKaC2)+2^(1+pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*5^(
pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-10^(5*pH+pKaA2)+10^(5*pH+pKaB2)-
10^(3*pH+pKaA1+pKaA2+pKaB2)+10^(3*pH+pKaA2+pKaB1+pKaB2)+10^(5*pH+pKaC2)
-
10^(3*pH+pKaA1+pKaA2+pKaC2)+10^(3*pH+pKaA2+pKaB2+pKaC2)+3*10^(3*pH+pKaB
1+pKaB2+pKaC2)+10^(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)+10^(3*pH+pKaA2+pKa
C1+pKaC2)+3*10^(3*pH+pKaB2+pKaC1+pKaC2)+10^(pH+pKaA1+pKaA2+pKaB2+pKaC1+
pKaC2)+3*10^(pH+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-
(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)
+100^pH)*(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zA+(10^(pH+pKaA2)+10^(
pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH)*(10^(pH+pK
aC2)+10^(pKaC1+pKaC2)+100^pH)*zB+10^(6*pH)*zC+10^(5*pH+pKaA2)*zC+10^(4*
pH+pKaA1+pKaA2)*zC+10^(5*pH+pKaB2)*zC+10^(4*pH+pKaA2+pKaB2)*zC+10^(3*pH
+pKaA1+pKaA2+pKaB2)*zC+10^(4*pH+pKaB1+pKaB2)*zC+10^(3*pH+pKaA2+pKaB1+pK
aB2)*zC+10^(2*pH+pKaA1+pKaA2+pKaB1+pKaB2)*zC+10^(5*pH+pKaC2)*zC+10^(4*p
H+pKaA2+pKaC2)*zC+10^(3*pH+pKaA1+pKaA2+pKaC2)*zC+10^(4*pH+pKaB2+pKaC2)*
zC+10^(3*pH+pKaA2+pKaB2+pKaC2)*zC+10^(2*pH+pKaA1+pKaA2+pKaB2+pKaC2)*zC+
10^(3*pH+pKaB1+pKaB2+pKaC2)*zC+10^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)*zC+10^
(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)*zC+10^(4*pH+pKaC1+pKaC2)*zC+10^(3*pH
+pKaA2+pKaC1+pKaC2)*zC+10^(2*pH+pKaA1+pKaA2+pKaC1+pKaC2)*zC+10^(3*pH+pK
aB2+pKaC1+pKaC2)*zC+10^(2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*zC+10^(pH+pKaA1+p
KaA2+pKaB2+pKaC1+pKaC2)*zC+10^(2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(pH+
pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKa
C2)*zC))
```

C.2.6.3 Ionic Strength

```
Ionic Strength = -((2^(-1-pH)*((-((-
(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH))*((-10^(-
pKw))*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKw)*(10^(pH+pKaI
ndicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH+pKaPhos3)
+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Cl+
10^(2*pH+pKw)*(2^(2+pH)*5^pH+10^pKaIndicator2)*(10^(2*pH+pKaPhos3)+10^(
pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Indicato
rTot+(10^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*((
10^pKw+100^pH)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPho
s1+pKaPhos2+pKaPhos3)+1000^pH)+10^(pH+pKw)*(10^(2*pH+pKaPhos3)+10^(pH+p
KaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Na+10^(2*pH+
pKw)*(2^(2+pH+pKaPhos3)*5^(pH+pKaPhos3)+10^(pKaPhos2+pKaPhos3)+9*100^pH
)*Phostot))-
10^pH*(10^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(
10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaP
hos3)+1000^pH)*(Ainit+Cinit)*(10^pKaA2*(2^(2+pKaA1)*5^pKaA1+10^pH)+2^(1
+pKaA2)*5^pKaA2*(2^(1+pKaA1)*5^pKaA1+10^pH)*zA+(10^(pH+pKaA2)+10^(pKaA1
+pKaA2)+100^pH)*zA^2))+10^pH*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(1
0^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH
```

```
+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+10
00^pH)*(Binit-
Cinit)*(10^pKaB2*(2^(2+pKaB1)*5^pKaB1+10^pH)+2^(1+pKaB2)*5^pKaB2*(2^(1+
pKaB1)*5^pKaB1+10^pH)*zB+(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH)*zB^2))
)*((-
2^(1+4*pH+pKaA1+pKaA2))*5^(4*pH+pKaA1+pKaA2)+2^(1+4*pH+pKaB1+pKaB2)*5^(
4*pH+pKaB1+pKaB2)+2^(1+4*pH+pKaB2+pKaC2)*5^(4*pH+pKaB2+pKaC2)+2^(1+2*pH
+pKaA2+pKaB1+pKaB2+pKaC2)*5^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)+2^(1+4*pH+pK
aC1+pKaC2)*5^(4*pH+pKaC1+pKaC2)+2^(1+2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*5^(2
*pH+pKaA2+pKaB2+pKaC1+pKaC2)+2^(2+2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*5^(2*pH
+pKaB1+pKaB2+pKaC1+pKaC2)+2^(1+pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*5^(
pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-10^(5*pH+pKaA2)+10^(5*pH+pKaB2)-
10^(3*pH+pKaA1+pKaA2+pKaB2)+10^(3*pH+pKaA2+pKaB1+pKaB2)+10^(5*pH+pKaC2)
-
10^(3*pH+pKaA1+pKaA2+pKaC2)+10^(3*pH+pKaA2+pKaB2+pKaC2)+3*10^(3*pH+pKaB
1+pKaB2+pKaC2)+10^(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)+10^(3*pH+pKaA2+pKa
C1+pKaC2)+3*10^(3*pH+pKaB2+pKaC1+pKaC2)+10^(pH+pKaA1+pKaA2+pKaB2+pKaC1+
pKaC2)+3*10^(pH+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-
(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)
+100^pH)*(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zA+(10^(pH+pKaA2)+10^(
pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH)*(10^(pH+pK
aC2)+10^(pKaC1+pKaC2)+100^pH)*zB+10^(6*pH)*zC+10^(5*pH+pKaA2)*zC+10^(4*
pH+pKaA1+pKaA2)*zC+10^(5*pH+pKaB2)*zC+10^(4*pH+pKaA2+pKaB2)*zC+10^(3*pH
+pKaA1+pKaA2+pKaB2)*zC+10^(4*pH+pKaB1+pKaB2)*zC+10^(3*pH+pKaA2+pKaB1+pK
aB2)*zC+10^(2*pH+pKaA1+pKaA2+pKaB1+pKaB2)*zC+10^(5*pH+pKaC2)*zC+10^(4*p
H+pKaA2+pKaC2)*zC+10^(3*pH+pKaA1+pKaA2+pKaC2)*zC+10^(4*pH+pKaB2+pKaC2)*
zC+10^(3*pH+pKaA2+pKaB2+pKaC2)*zC+10^(2*pH+pKaA1+pKaA2+pKaB2+pKaC2)*zC+
10^(3*pH+pKaB1+pKaB2+pKaC2)*zC+10^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)*zC+10^
(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)*zC+10^(4*pH+pKaC1+pKaC2)*zC+10^(3*pH
+pKaA2+pKaC1+pKaC2)*zC+10^(2*pH+pKaA1+pKaA2+pKaC1+pKaC2)*zC+10^(3*pH+pK
aB2+pKaC1+pKaC2)*zC+10^(2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*zC+10^(pH+pKaA1+p
KaA2+pKaB2+pKaC1+pKaC2)*zC+10^(2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(pH+
pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKa
C2)*zC)-((-(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH))*((-10^(-
pKw))*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKw)*(10^(pH+pKaI
ndicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH+pKaPhos3)
+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Cl+
10^(2*pH+pKw)*(2^(1+pH)*5^pH+10^pKaIndicator2)*(10^(2*pH+pKaPhos3)+10^(
pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*Indicato
rTot+(10^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*((
-
10^pKw+100^pH)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPho
s1+pKaPhos2+pKaPhos3)+1000^pH)-
10^(pH+pKw)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+
pKaPhos2+pKaPhos3)+1000^pH)*Na+10^(2*pH+pKw)*(2^(1+pH+pKaPhos3)*5^(pH+p
KaPhos3)+10^(pKaPhos2+pKaPhos3)+3*100^pH)*Phostot))+10^pH*(10^(pH+pKaIn
dicator2)+10^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH+pKaPhos3)+
10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*(Ain
it+Cinit)*(10^pKaA2*(2^(1+pKaA1)*5^pKaA1+10^pH)+(10^(pH+pKaA2)+10^(pKaA
1+pKaA2)+100^pH)*zA))-
10^pH*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaIndicator2)+10
^(pKaIndicator1+pKaIndicator2)+100^pH)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPh
os2+pKaPhos3)+10^(pKaPhos1+pKaPhos2+pKaPhos3)+1000^pH)*(Binit-
Cinit)*(10^pKaB2*(2^(1+pKaB1)*5^pKaB1+10^pH)+(10^(pH+pKaB2)+10^(pKaB1+p
```

```
KaB2)+100^pH)*zB))*((-
2^(2+4*pH+pKaA1+pKaA2))*5^(4*pH+pKaA1+pKaA2)+2^(2+4*pH+pKaB1+pKaB2)*5^(
4*pH+pKaB1+pKaB2)+2^(1+4*pH+pKaB2+pKaC2)*5^(4*pH+pKaB2+pKaC2)-
2^(1+2*pH+pKaA1+pKaA2+pKaB2+pKaC2)*5^(2*pH+pKaA1+pKaA2+pKaB2+pKaC2)+2^(
3*pH+pKaB1+pKaB2+pKaC2)*5^(1+3*pH+pKaB1+pKaB2+pKaC2)+2^(2+2*pH+pKaA2+pK
aB1+pKaB2+pKaC2)*5^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)+2^(2+4*pH+pKaC1+pKaC2
)*5^(4*pH+pKaC1+pKaC2)+2^(3*pH+pKaB2+pKaC1+pKaC2)*5^(1+3*pH+pKaB2+pKaC1
+pKaC2)+2^(2+2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*5^(2*pH+pKaA2+pKaB2+pKaC1+pK
aC2)+2^(3+2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*5^(2*pH+pKaB1+pKaB2+pKaC1+pKaC2
)+2^(2+pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*5^(pKaA1+pKaA2+pKaB1+pKaB2+
pKaC1+pKaC2)-10^(5*pH+pKaA2)+10^(5*pH+pKaB2)-
3*10^(3*pH+pKaA1+pKaA2+pKaB2)+3*10^(3*pH+pKaA2+pKaB1+pKaB2)+10^(5*pH+pK
aC2)-
3*10^(3*pH+pKaA1+pKaA2+pKaC2)+10^(3*pH+pKaA2+pKaB2+pKaC2)+10^(pH+pKaA1+
pKaA2+pKaB1+pKaB2+pKaC2)+3*10^(3*pH+pKaA2+pKaC1+pKaC2)+10^(pH+pKaA1+pKa
A2+pKaB2+pKaC1+pKaC2)+7*10^(pH+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-
2^(1+pKaA2)*5^pKaA2*(2^(1+pKaA1)*5^pKaA1+10^pH)*(10^(pH+pKaB2)+10^(pKaB
1+pKaB2)+100^pH)*(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zA-
(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)
+100^pH)*(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zA^2+2^(1+pKaB2)*5^pKa
B2*(2^(1+pKaB1)*5^pKaB1+10^pH)*(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*
(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zB+(10^(pH+pKaA2)+10^(pKaA1+pKa
A2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH)*(10^(pH+pKaC2)+10^(
pKaC1+pKaC2)+100^pH)*zB^2+2^(1+5*pH+pKaC2)*5^(5*pH+pKaC2)*zC+2^(1+4*pH+
pKaA2+pKaC2)*5^(4*pH+pKaA2+pKaC2)*zC+2^(1+3*pH+pKaA1+pKaA2+pKaC2)*5^(3*
pH+pKaA1+pKaA2+pKaC2)*zC+2^(1+4*pH+pKaB2+pKaC2)*5^(4*pH+pKaB2+pKaC2)*zC
+2^(1+3*pH+pKaA2+pKaB2+pKaC2)*5^(3*pH+pKaA2+pKaB2+pKaC2)*zC+2^(1+2*pH+p
KaA1+pKaA2+pKaB2+pKaC2)*5^(2*pH+pKaA1+pKaA2+pKaB2+pKaC2)*zC+2^(1+3*pH+p
KaB1+pKaB2+pKaC2)*5^(3*pH+pKaB1+pKaB2+pKaC2)*zC+2^(1+2*pH+pKaA2+pKaB1+p
KaB2+pKaC2)*5^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)*zC+2^(1+pH+pKaA1+pKaA2+pKa
B1+pKaB2+pKaC2)*5^(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)*zC+2^(2+4*pH+pKaC1
+pKaC2)*5^(4*pH+pKaC1+pKaC2)*zC+2^(2+3*pH+pKaA2+pKaC1+pKaC2)*5^(3*pH+pK
aA2+pKaC1+pKaC2)*zC+2^(2+2*pH+pKaA1+pKaA2+pKaC1+pKaC2)*5^(2*pH+pKaA1+pK
aA2+pKaC1+pKaC2)*zC+2^(2+3*pH+pKaB2+pKaC1+pKaC2)*5^(3*pH+pKaB2+pKaC1+pK
aC2)*zC+2^(2+2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*5^(2*pH+pKaA2+pKaB2+pKaC1+pK
aC2)*zC+2^(2+pH+pKaA1+pKaA2+pKaB2+pKaC1+pKaC2)*5^(pH+pKaA1+pKaA2+pKaB2+
pKaC1+pKaC2)*zC+2^(2+2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*5^(2*pH+pKaB1+pKaB2+
pKaC1+pKaC2)*zC+2^(2+pH+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*5^(pH+pKaA2+pKaB
1+pKaB2+pKaC1+pKaC2)*zC+2^(2+pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*5^(pK
aA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(6*pH)*zC^2+10^(5*pH+pKaA2)*zC
^2+10^(4*pH+pKaA1+pKaA2)*zC^2+10^(5*pH+pKaB2)*zC^2+10^(4*pH+pKaA2+pKaB2
)*zC^2+10^(3*pH+pKaA1+pKaA2+pKaB2)*zC^2+10^(4*pH+pKaB1+pKaB2)*zC^2+10^(
3*pH+pKaA2+pKaB1+pKaB2)*zC^2+10^(2*pH+pKaA1+pKaA2+pKaB1+pKaB2)*zC^2+10^
(5*pH+pKaC2)*zC^2+10^(4*pH+pKaA2+pKaC2)*zC^2+10^(3*pH+pKaA1+pKaA2+pKaC2
)*zC^2+10^(4*pH+pKaB2+pKaC2)*zC^2+10^(3*pH+pKaA2+pKaB2+pKaC2)*zC^2+10^(
2*pH+pKaA1+pKaA2+pKaB2+pKaC2)*zC^2+10^(3*pH+pKaB1+pKaB2+pKaC2)*zC^2+10^
(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)*zC^2+10^(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC
2)*zC^2+10^(4*pH+pKaC1+pKaC2)*zC^2+10^(3*pH+pKaA2+pKaC1+pKaC2)*zC^2+10^
(2*pH+pKaA1+pKaA2+pKaC1+pKaC2)*zC^2+10^(3*pH+pKaB2+pKaC1+pKaC2)*zC^2+10
^(2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*zC^2+10^(pH+pKaA1+pKaA2+pKaB2+pKaC1+pKa
C2)*zC^2+10^(2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*zC^2+10^(pH+pKaA2+pKaB1+pKaB
2+pKaC1+pKaC2)*zC^2+10^(pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*zC^2)))/5^
pH/((10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pK
aB2)+100^pH)*(10^(pH+pKaIndicator2)+10^(pKaIndicator1+pKaIndicator2)+10
```

```
0^pH)*(10^(2*pH+pKaPhos3)+10^(pH+pKaPhos2+pKaPhos3)+10^(pKaPhos1+pKaPho
s2+pKaPhos3)+1000^pH)*((-
2^(1+4*pH+pKaA1+pKaA2))*5^(4*pH+pKaA1+pKaA2)+2^(1+4*pH+pKaB1+pKaB2)*5^(
4*pH+pKaB1+pKaB2)+2^(1+4*pH+pKaB2+pKaC2)*5^(4*pH+pKaB2+pKaC2)+2^(1+2*pH
+pKaA2+pKaB1+pKaB2+pKaC2)*5^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)+2^(1+4*pH+pK
aC1+pKaC2)*5^(4*pH+pKaC1+pKaC2)+2^(1+2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*5^(2
*pH+pKaA2+pKaB2+pKaC1+pKaC2)+2^(2+2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*5^(2*pH
+pKaB1+pKaB2+pKaC1+pKaC2)+2^(1+pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*5^(
pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-10^(5*pH+pKaA2)+10^(5*pH+pKaB2)-
10^(3*pH+pKaA1+pKaA2+pKaB2)+10^(3*pH+pKaA2+pKaB1+pKaB2)+10^(5*pH+pKaC2)
-
10^(3*pH+pKaA1+pKaA2+pKaC2)+10^(3*pH+pKaA2+pKaB2+pKaC2)+3*10^(3*pH+pKaB
1+pKaB2+pKaC2)+10^(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)+10^(3*pH+pKaA2+pKa
C1+pKaC2)+3*10^(3*pH+pKaB2+pKaC1+pKaC2)+10^(pH+pKaA1+pKaA2+pKaB2+pKaC1+
pKaC2)+3*10^(pH+pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)-
(10^(pH+pKaA2)+10^(pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)
+100^pH)*(10^(pH+pKaC2)+10^(pKaC1+pKaC2)+100^pH)*zA+(10^(pH+pKaA2)+10^(
pKaA1+pKaA2)+100^pH)*(10^(pH+pKaB2)+10^(pKaB1+pKaB2)+100^pH)*(10^(pH+pK
aC2)+10^(pKaC1+pKaC2)+100^pH)*zB+10^(6*pH)*zC+10^(5*pH+pKaA2)*zC+10^(4*
pH+pKaA1+pKaA2)*zC+10^(5*pH+pKaB2)*zC+10^(4*pH+pKaA2+pKaB2)*zC+10^(3*pH
+pKaA1+pKaA2+pKaB2)*zC+10^(4*pH+pKaB1+pKaB2)*zC+10^(3*pH+pKaA2+pKaB1+pK
aB2)*zC+10^(2*pH+pKaA1+pKaA2+pKaB1+pKaB2)*zC+10^(5*pH+pKaC2)*zC+10^(4*p
H+pKaA2+pKaC2)*zC+10^(3*pH+pKaA1+pKaA2+pKaC2)*zC+10^(4*pH+pKaB2+pKaC2)*
zC+10^(3*pH+pKaA2+pKaB2+pKaC2)*zC+10^(2*pH+pKaA1+pKaA2+pKaB2+pKaC2)*zC+
10^(3*pH+pKaB1+pKaB2+pKaC2)*zC+10^(2*pH+pKaA2+pKaB1+pKaB2+pKaC2)*zC+10^
(pH+pKaA1+pKaA2+pKaB1+pKaB2+pKaC2)*zC+10^(4*pH+pKaC1+pKaC2)*zC+10^(3*pH
+pKaA2+pKaC1+pKaC2)*zC+10^(2*pH+pKaA1+pKaA2+pKaC1+pKaC2)*zC+10^(3*pH+pK
aB2+pKaC1+pKaC2)*zC+10^(2*pH+pKaA2+pKaB2+pKaC1+pKaC2)*zC+10^(pH+pKaA1+p
KaA2+pKaB2+pKaC1+pKaC2)*zC+10^(2*pH+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(pH+
pKaA2+pKaB1+pKaB2+pKaC1+pKaC2)*zC+10^(pKaA1+pKaA2+pKaB1+pKaB2+pKaC1+pKa
C2)*zC)))
```

C.2.7 MATLAB Script Explanation

The analytical expressions for conversion and absorbance are coded into a MATLAB framework for evaluation. Since this framework includes sensitivity analysis and evaluation across a range of pH values, care is taken to replace all instances of matrix-wise operators (`*`, `/`, `^`) with their element-wise counterparts, (`.*`, `./`, `.^`). Fortunately, this can be done using a 'Find and Replace' command in most text editors.

For convenience of predicting assay response curves for many different reactions and reaction conditions, the design of an assay experiment and necessary model parameters can be specified in a Microsoft Excel spreadsheet. The filename of this spreadsheet is specified in the MATLAB code, and parameters are pulled from the spreadsheet into a Struct, "`Assay`" when the code is run. A new Struct, "`ModelOut`" is then generated which has fields for each model input, as well as for data that are generated during the computation.

The MATLAB framework consists of a several primary *functions*, which are explained here:

1. Function `main()`
    a. Initializes the MATLAB workspace.
    b. User can adjust the pH range and resolution over which to evaluate the model, and the number of Monte Carlo samples to pull from the parameter space (if desired).

c. NOTE: The input pH range is concentration-based (pH_c), not activity-based (pH_a).

d. Prints to the command window the number (corresponding to the column in the Excel spreadsheet), and name of the assay being evaluated.

e. Calls `func_DefineAssay(ExcelFile, MC_samps, pH_samps)` once in order to initialize the main struct which organizes the model data.

f. Calls `func_RunSubAssay(SubAssay,pH)` once for each set of assay conditions described in the Excel spreadsheet.

g. This function is also a convenient place to produce a plot displaying multiple assay runs. This may be desired if the user wishes to compare multiple assays, such as different conditions for evaluating the same reaction. This is also a convenient place to load and plot experimental data for analysis or comparison to the modeled assay response curves.

2. Function `func_DefineAssay(ExcelFile, MC_samps, pH_samps)`

a. Reads Assay conditions and parameters in from specified Excel spreadsheet, and initializes the main Struct which organizes the model data.

b. If the user has specified to run Monte Carlo sensitivity/error analysis, this function pulls parameters from a normal distribution with the mean and standard deviation for that parameter, as specified in the Excel spreadsheet.

3. Function `func_RunSubAssay(SubAssay, pH)`

a. Generates a 3-dimensional matrix of inputs which are fed into `func_Calculations(SubAssay.ParsIn, pH)`. By MATLAB notation, the matrix is addressed by (row, column, page)

   i. Rows (each column x page slice): The first row of the matrix contains the nominal (mean) values of all input parameters, and repeats them across for each pH value in the specified range. Each subsequent row contains set of parameters from the Monte Carlo sampling of the parameter space, if specified, also repeated across all pH values.

   ii. Columns (each row x page slice): Each column contains the same data (nominal and Monte Carlo values for each parameter) repeated. There are as many columns as there are specified pH values.

   iii. Pages (each row x column slice): Each page corresponds to a single one of the model input parameters.

b. Prints to the command window the valid ranges for pH, conversion, absorbance, and ionic strength for the given assay number.

c. If no Monte Carlo analysis is performed, three plots are produced here for each assay run:

   i. An XYY plot, with activity-based pH, pH_a, on the horizontal axis, reaction conversion on the left vertical axis, and absorbance on the right vertical axis. Recall

that pH is the independent variable for the expressions of both conversion and absorbance.

ii. An XY plot, with absorbance on the horizontal axis and conversion on the vertical axis. From this plot, the user may determine what reaction conversion is achieved for a particular absorbance or absorbance ratio.

iii. An XY plot, with activity-based pH on the horizontal axis and ionic strength on the vertical axis.

d. If Monte Carlo analysis *is* performed, either to determine sensitivity of the model to variation in a single parameter, or to estimate error in the model due to uncertainty in multiple parameters, then several additional things happen. (Monte Carlo analysis is discussed in more detail in C.2.9, Model Resolution, Sensitivity, and Error Analysis.

i. The values of conversion and absorbance at each pH are fitted to a truncated normal distribution centered on the nominal values, over the entire pH range where conversion is between zero and one, or where absorbance is greater than zero. Examples of the frequency distribution of conversion at three values of pH are shown.

ii. The standard deviation bands are appended to the XYY plot showing conversion and absorbance vs. pH.

iii. The values of conversion and absorbance plotted against each other (parametric with respect to pH) are fitted to a truncated normal distribution, over the entire pH

range where conversion is between zero and one and absorbance is greater than zero. Examples of the frequency distribution of conversion at three values of pH are shown. This is reflective of the uncertainty incurred when using a measured absorbance value (or absorbance ratio) to determine reaction conversion.

4. Function `func_Calculations(SubAssay.ParsIn, pH)`

   a. This function is the mathematical heart of the model, and calls the algebraic expressions (from Mathematica) for ionic strength, conversion and absorbance as functions of pH (the concentration based pH, pH_c). All multiplication, division, and exponentiation is element-wise, so that the expressions can be evaluated over a range of pH values and Monte Carlo parameter samples simultaneously.

   b. First, this function performs three iterations of ionic strength calculation and pKa adjustment by calling `func_IonicStrength(ParsIn_New, pH_conc)` and `func_pKas(ParsIn, IonicStrength)`. The results of each iteration are appended to Figure 1, which is cleared for each new sub assay.

   c. `func_Conversion(ParsIn_New, pH_conc)` is called, which returns the reaction conversion required to achieve each input pH.

   d. `func_Absorbance(ParsIn_New, pH_conc)` is called, which returns the reaction conversion required to achieve each input pH.

   e. The concentration-based pH, pH_c, is then adjusted for ionic strength using the extended Debye-Hückel correction, resulting in the activity-based pH, pH_a. The

activity-based pH corresponds to the value that would be read from a typical potentiometric pH meter. This is the pH that is then returned to the user and displayed in plots.

f. A 3-dimensional matrix is constructed, with one column corresponding to each input activity-based pH, one row corresponding to the nominal parameter values and one additional row for each set of Monte Carlo parameters, and one page each containing values of the activity-based pH, the conversion, the absorbance, and the ionic strength under each combination of the aforementioned input parameters.

g. The matrix is then trimmed to the range of pH values for which the nominal conversion is between zero and one. For this reason, the pH range displayed in graphs that are produced may not be as broad as that initially specified.

5. Functions `func_Conversion(ParsIn, pH)`, `func_Absorbance(ParsIn, pH)`, and `func_IonicStrength(ParsIn, pH)`

   a. Evaluate the algebraic Mathematica expressions for conversion, absorbance, and ionic strength, respectively.

   b. Each of these functions takes the concentration-based pH, pH_c, as an input parameter.

6. Function `func_pKas(ParsIn, IonicStrength)`

   a. Performs the extended Debye-Hückel correction to convert thermodynamic pKa to a concentration-basis.

At the end of the calculation, everything is contained within a struct called `ModelOut`. The hierarchy and contents of that Struct are provided here:

1. `ModelOut`: The top level of the Struct. It has one index for each column (each specified set of assay conditions) in the Excel Spreadsheet.
    a. `ModelOut(i)`: The index corresponding to the desired set of assay conditions.
        i. `A`: The name of the reaction substrate, A.
        ii. `B`: The name of the first product, B.
        iii. `C`: The name of the second product, C.
        iv. `Title`: The title of the assay.
        v. `Titlestr`: The number and title of the assay, for use in graphs.
        vi. `Input`: Container for arrays of [Nominal value; Monte Carlo Values] for each model parameter
            1. `pKaA1`: First pK_a^{th} of the reactant, A.
            2. `pKaA2`: Second pK_a^{th} of the reactant, A.
            3. `zA`: Charge number of the least protonated state of the reactant, A.
            4. `pKaB1`: First pK_a^{th} of the first product, B.

5. `pKaB2`: Second pK_a^{th} of the first product, B.

6. `zB`: Charge number of the least protonated state of the first product, B.

7. `pKaC1`: First pK_a^{th} of the second product, C.

8. `pKaC2`: Second pK_a^{th} of the second product, C.

9. `zC`: Charge number of the least protonated state of the second product, C.

10. `pKaIndicator1`: First pK_a^{th} of the colorimetric pH indicator.

11. `pKaIndicator2`: Second pK_a^{th} of the colorimetric pH indicator.

12. `pKaPhos1`: First pK_a^{th} of phosphoric acid.

13. `pKaPhos2`: Second pK_a^{th} of phosphoric acid.

14. `pKaPhos3`: Third pK_a^{th} of phosphoric acid.

15. `pKw`: Self-Ionization constant of water (at zero ionic strength)

16. `Ainit`: Initial concentration of the reactant, A.

17. `Binit`: Initial concentration of the first product, B.

18. `Cinit`: Initial concentration of the second product, C.

19. `Phostot`: Enduring analytical concentration of the phosphate buffer.

20. `Na`: Enduring analytical concentration of the sodium ion.

21. `Cl`: Enduring analytical concentration of the chloride ion.

22. `IndicatorTot`: Enduring analytical concentration of the colorimetric indicator.

23. `AbsIndH2`: Absorbance, or absorbance ratio of the most protonated state of the indicator.

24. `AbsIndH`: Absorbance, or absorbance ratio of the intermediately protonated state of the indicator.

25. `AbsInd`: Absorbance, or absorbance ratio of the least protonated state of the indicator.

vii. `ParsIn`: A 3-dimensional matrix of values that is used by `func_Calculations()`. It has one row corresponding to the nominal input parameters, and one additional row corresponding to each set of Monte Carlo parameter values. It has one column for each input concentration-based pH value, and one page for each of the 25 input parameter fields.

viii. `ParsOut.Raw`: A 3-dimensional matrix of values that are returned by `func_Calculations()`. The matrix has one column corresponding to each input activity-based pH, one row corresponding to the nominal parameter values and one additional row for each set of Monte Carlo parameters, and one page each

containing values of the activity-based pH, the conversion, the absorbance, and the ionic strength under each combination of the input parameters.

ix. `Output`: This contains the useful calculated values that are the primary result of the model.

1. `pH.Nominal`: A single array containing all of the activity-based pH values, pH_a.

2. `X`: Container for all of the arrays of values relating to conversion.

   a. `Nominal`: The conversion required to produce each of the nominal pH_a values, corresponding to the nominal values of the input parameters.

   b. `Mean_fit`: The fitted mean of the calculated conversion values, aligned with respect to pH_a, if Monte Carlo analysis is performed.

   c. `Std_fit`: The fitted standard deviation of conversion with respect to pH_a.

   d. `High_fit`: Nominal + 1 standard deviation, for conversion with respect to pH_a.

   e. `Low_fit`: Nominal - 1 standard deviation, for conversion with respect to pH_a.

   f. `Mean_fit_Abs`: The fitted mean of conversion with respect to absorbance (or absorbance ratio).

   g. `Std_fit_Abs`: The fitted standard deviation of conversion with respect to absorbance (or absorbance ratio).

h. `High_fit_Abs`: Mean + 1 standard deviation for conversion with respect to absorbance (or absorbance ratio).

i. `Low_fit_Abs`: Mean - 1 standard deviation for conversion with respect to absorbance (or absorbance ratio).

3. `Absorbance`: Container for the arrays relating to absorbance or absorbance ratio.

   a. `Nominal`: The calculated absorbance, of the solution at each of the nominal pH_a values, corresponding to the nominal values of the input parameters.

   b. `Mean_fit`: The fitted mean of the calculated absorbance values, aligned with respect to pH_a, if Monte Carlo analysis is performed.

   c. `Std_fit`: The fitted standard deviation of absorbance with respect to pH_a.

   d. `High_fit`: Nominal + 1 standard deviation, for absorbance with respect to pH_a.

   e. `Low_fit`: Nominal - 1 standard deviation, for absorbance with respect to pH_a.

4. `IonicStrength.Nominal`: The calculated ionic strengths corresponding to the nominal pH_a values.

C.2.8 Modifying the Model: Making Changes to the System of Equations

If the user should wish to modify the model beyond the scope of merely adjusting input parameters in the Excel spreadsheet, or plotting options in MATLAB, they may do so.

It is a relatively straightforward matter to alter the system of equations that feeds into the model, and propagate that change through into MATLAB. This procedure may require intermediate proficiency in MATLAB and Wolfram Mathematica.

1. Copy the Mathematica Input from C.2.5, Mathematica Input: Solving the Model Equations into a new notebook in Wolfram Mathematica.

2. Modify the system of equations as desired and evaluate the notebook. It may take some time. It may not be feasible to solve many systems for pH as the dependent variable.

3. Right click on the output for conversion (X) and in the context menu select ‘Convert To’ and then ‘InputForm’.

4. Copy the output into any text editor and perform the following changes using the ‘Find and Replace’ command to make the replacements shown in Table 28:

Table 28: 'Find and Replace' guide for transferring equations from Mathematica to MATLAB.

Find	Replace With
“*”	“.*”
“/”	“./”
“^”	“.^”
[space]	[nothing]
[tab]	[nothing]
[newline]	[nothing]

5. Copy the resulting text into MATLAB, replacing the existing expression for conversion. Absorbance, and ionic strength in their respective functions.

6. Repeat steps 3-5 for the expression for absorbance (or absorbance ratio) and for ionic strength.

7. Adjust the Excel Spreadsheet and structure handling routines in the model if necessary (i.e., if parameters have been added to or removed from the model.
8. Run the modified model.

C.2.9 Model Resolution, Sensitivity, and Error Analysis

The core of the model is designed to permit calculation of conversion and absorption from pH using large arrays of alternate values of input parameters simultaneously. Model sensitivity to variations in a particular parameter, or prediction error due to uncertainty in the true value of multiple parameters can therefore be estimated by plugging entire arrays of pre-selected parameter sets into the expressions for absorbance and conversion.

Before running the model, the user must set the values of `MC_samps` and `pH_samps` in the `main()` function. The functions of these two variables are described presently:

`MC_samps`**:**

Sets the number of Monte Carlo samples that are generated for each parameter. These samples are random numbers generated from a normal distribution with the mean and standard deviation specified in the Excel Spreadsheet. To evaluate sensitivity to changes in only one parameter, all other parameters should have their standard deviation set to zero. The model will still generate `MC_samps` samples for all parameters, but as they will come from a distribution with zero variance, they will all be exactly the nominal value.

`pH_samps`**:**

Sets the number of linearly-spaced points within the pH range (referring to concentration-based pH, pH_c). Since the calculated conversion will only be within the physically meaningful range of 0 to 1 for a small subset of these pH values, it is useful to contract the range or increase the value of `pH_samps` in order to achieve sufficient resolution for a satisfactory calculation. If the model is not able to get sufficiently close to conversions of 0 or 1 and appears to be cut off within the middle of this range, increasing the resolution would be a good first-pass troubleshooting approach.

The function `func_Calculations(ParsIn, pH_conc)`, which calls the functions containing expressions for ionic strength, absorbance, and conversion requires two arguments: `ParsIn` and `pH_conc`. The architectures of the arrays for `ParsIn` and `pH_conc` are shown in Figure 69. ParsIn is a 3-dimensional matrix of parameter values, and `pH_conc` is a 2-dimensional matrix of concentration-based pH values. `ParsIn` has one page for every parameter, one row for the nominal parameter values, and `MC_samps` number of additional rows for the sampled parameter values. This row-page slice is

repeated across pH_samps number of columns, providing an input set of parameters for each input pH value (so each column of ParsIn is identical). pH_conc has one column for each input pH value, and thus like ParsIn has pH_samps number of columns. Also like ParsIn, pH has MC_samps + 1 number of rows. Each row of pH_conc is identical. When no Monte Carlo analysis is performed, and the value of MC_samps is set to zero, then both ParsIn and pH_conc are reduced to a single row, and only the nominal parameter values specified in the Excel spreadsheet are used. Finally, within func_Assay(), each parameter is pulled as a 2-dimensional array from its corresponding page in ParsIn. Therefore, when the model expressions are evaluated, all inputs are of size (MC_samps + 1, pH_samps), and the element-wise operators enable the calculation of ionic strength, conversion, and absorbance to be determined at all pH values with all Monte Carlo parameter sample sets simultaneously.

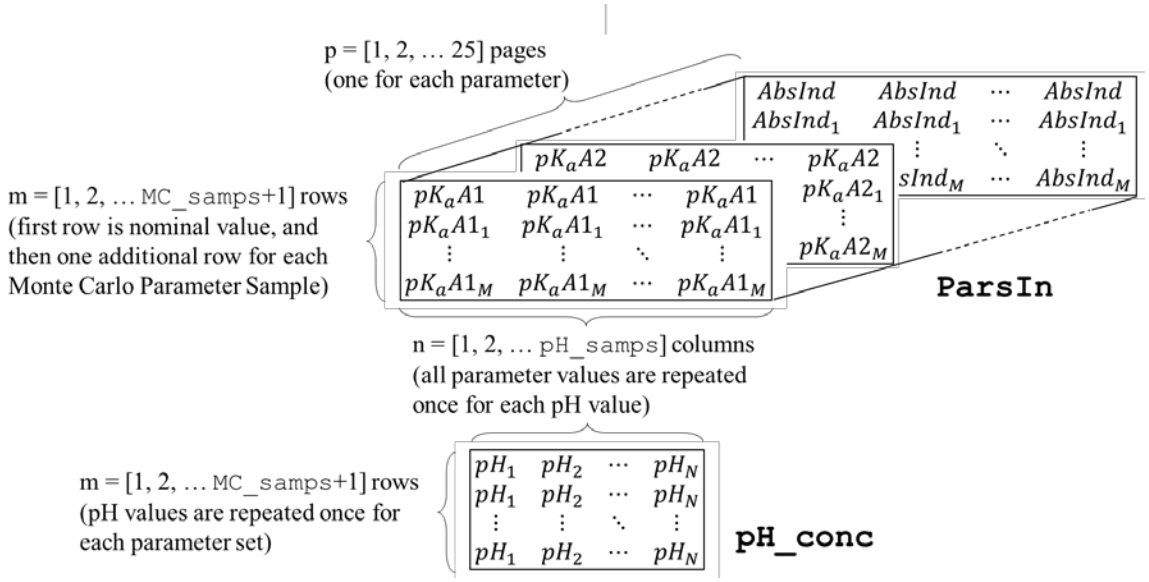


Figure 69: Architecture of ParsIn and pH_conc, the two inputs of func_Assay().

When Monte Carlo sampling is performed (i.e., when MC_samps is nonzero), each input concentration-based pH value has one corresponding absorbance, conversion level, and activity-based pH for each set of input parameters (MC_samps + 1 values, including the nominal value). Although each column in ParsOut corresponds to the same input pH_c, changes in ionic strength due to the different input conditions may result in differing values of pH_a in a given column for each Monte Carlo parameter set. Therefore, to estimate the uncertainty in the calculated value resulting from the uncertainty in the input Monte Carlo parameter set, at each activity-based pH, each data series is interpolated to

produce data that is aligned by pH_a, rather than by pH_c. In the case of estimating uncertainty in the direct relationship between conversion and absorbance, a similar strategy is used to align the calculated conversion data series by the calculated absorbance values. In either case, it is necessary to trim values which are nonphysical (i.e., conversions greater than 100 %). Since the datasets are trimmed, simply 'calculating' their mean and standard deviation would provide incorrect values (for example, the mean would be shifted away from the trim cutoff). Therefore, each aligned dataset is fitted with a truncated normal distribution, providing a fitted estimate of the mean and standard deviation of each ordinate dataset at each location on the abscissa.

An example of the type of output a user might expect to see from this type of analysis is shown in Figure 70. The main plots in Panels A and B both show the relationships between Conversion (y-axis) and Absorbance Ratio (x-axis) for the conversion of 100 mM methyl picolinate with $[Na^+] = 0.180 \pm 0.009$ M (a 5 % deviation from the nominal value). Both panels show the estimates of uncertainty in conversion relative to observed absorbance, however the calculation displayed in Panel A used 50 Monte Carlo samples, while that in Panel B used 10,000. In each panel, the subplots show the frequency of conversion values at select locations on the x-axis, along with the normal distribution that was used to determine the mean and standard deviation. At the extremes of absorbance ratio, the data (and thus the fitted distributions) are truncated at zero or one, corresponding to the physically meaningful limits of conversion. If insufficient Monte Carlo samples are used, the data will not appear to be smooth. Either insufficient Monte Carlo samples or significant deviations in pH_a between Monte Carlo sets may cause the mean value curve not to overlay with the curve representing the nominal calculation.

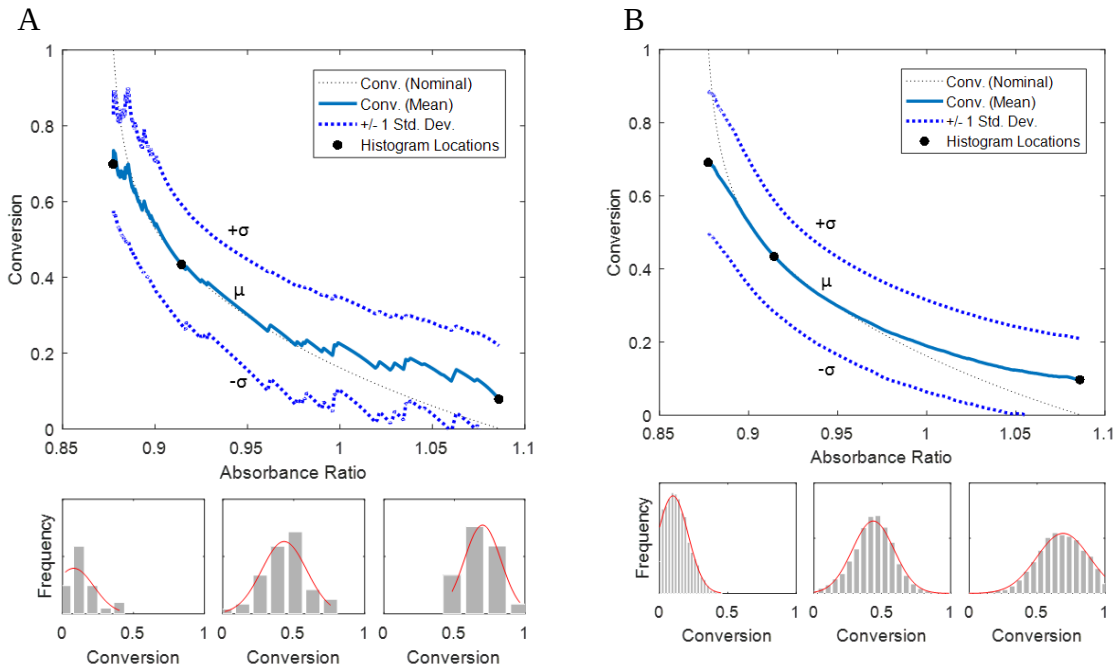


Figure 70: Examples of colorimetric assay sensitivity analysis using Monte Carlo simulation. The nominal conversion of 100 mM methyl picolinate vs pH is plotted against absorbance ratio in Panels A and B. Analysis of sensitivity of the assay to [Na^+] is performed by running the model with values of [Na^+] sampled from a normal distribution with $\mu = 0.180$ M and $\sigma = 0.009$ M. Panel A shows the result using 50 sampled values of [Na^+], while Panel B shows the result after 10,000 samples. The fitted and mean is plotted in addition to the nominal curve and bands are shown at ± 1 standard deviation. Subpanels display histograms (grey) showing the distributions of calculated conversions at select values of absorbance ratio, overlaid with the fitted probability density functions (red) corresponding to the fitted mean and standard deviations at those values of absorbance ratio.

APPENDIX D. PERMISSION FOR USE OF COPYRIGHTED MATERIALS

D.1 Furfural Degradation Pathway Map

The author has been granted permission for use of the furfural degradation pathway map image (Figure 11) by email on Thursday, February 15, 2018. The text of the email is copied below:

RE: [KEGG Copyright Permission] 180036

Dear Harrison Bellow Rose,

Thank you for contacting us for copyright permission of KEGG.

Permission is granted to you to publish the following KEGG pathway map image
in your doctoral thesis "Toward an Enzymatic Route to 2,5-Furandicarboxylic Acid"
written by Harrison Bellow Rose:
- Furfural degradation (map00365)

subject to the condition that the original source is properly acknowledged.

Please see more details in http://www.kegg.jp/kegg/kegg1.html.

Sincerely,
Miwako Matsumoto
Kanehisa Laboratories

REFERENCES

[1] Blunden, J.; Arndt, D. S., State of the Climate in 2013. *Bull. Am. Meteorol. Soc.* **2014,** *95* (7), S1-S257.

[2] Arndt, D. S.; Blunden, J.; Dunn, R. J. H., State of the Climate in 2016. *Bull. Am. Meteorol. Soc.* **2017,** *98* (8), Si-S277.

[3] Boot-Handford, M. E.; Abanades, J. C.; Anthony, E. J.; Blunt, M. J.; Brandani, S.; Mac Dowell, N.; Fernandez, J. R.; Ferrari, M. C.; Gross, R.; Hallett, J. P.; Haszeldine, R. S.; Heptonstall, P.; Lyngfelt, A.; Makuch, Z.; Mangano, E.; Porter, R. T. J.; Pourkashanian, M.; Rochelle, G. T.; Shah, N.; Yao, J. G.; Fennell, P. S., Carbon capture and storage update. *Energy Environ. Sci.* **2014,** *7* (1), 130-189.

[4] Bozell, J. J.; Petersen, G. R., Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's "Top 10" revisited. *Green Chem.* **2010,** *12* (4), 539-554.

[5] *The Future of PET Packating to 2021*. Smithers Pira: 2018.

[6] Köpnick, H.; Schmidt, M.; AG, B.; Brügging, W.; Rüter, J.; Aktiengesellschaft, H.; Kaminsky, W., Polyesters. In *Ullmann's Encyclopedia of Industrial Chemistry*, 5th, completely rev. ed.; Gerhartz, W.; Yamamoto, Y. S.; Campbell, F. T.; Pfefferkorn, R.; Rounsaville, J. F.; Ullmann, F., Eds. VCH: Weinheim, Federal Republic of Germany; Deerfield Beach, FL, USA, 1985; Vol. A21, pp 227-251.

[7] Kriegel, R. M., Georgia Institute of Technology Society of Plastics Engineers Invited Industrial Seminar. Atlanta, Georgia, United States, 2014.

[8] Gandini, A.; Silvestre, A. J. D.; Neto, C. P.; Sousa, A. F.; Gomes, M., The Furan Counterpart of Poly(ethylene terephthalate): An Alternative Material Based on Renewable Resources. *Journal of Polymer Science, Part A: Polymer Chemistry* **2009,** *47* (1), 295-298.

[9] Moreau, C.; Belgacem, M. N.; Gandini, A., Recent catalytic advances in the chemistry of substituted furans from carbohydrates and in the ensuing polymers. *Top. Catal.* **2004,** *27* (1-4), 11-30.

[10] Burgess, S. K.; Leisen, J. E.; Kraftschik, B. E.; Mubarak, C. R.; Kriegel, R. M.; Koros, W. J., Chain Mobility, Thermal, and Mechanical Properties of Poly(ethylene furanoate) Compared to Poly(ethylene terephthalate). *Macromolecules* **2014,** *47* (4), 1383-1391.

[11] Burgess, S. K.; Kriegel, R. M.; Koros, W. J., Carbon Dioxide Sorption and Transport in Amorphous Poly(ethylene furanoate). *Macromolecules* **2015,** *48* (7), 2184-2193.

[12] de Jong, E.; Dam, M. A.; Sipos, L.; Gruter, G. J. M., Furandicarboxylic Acid (FDCA), A Versatile Building Block for a Very Interesting Class of Polyesters. **2012,** *1105*, 1-13.

[13] De Jong, E.; Kemeling, N. In *PEF: Opportunities for a novel 100% bio-based polyester*, PETnology, Nuremberg, Germany, November 9, 2015; Avantium: Nuremberg, Germany, 2015.

[14] Werpy, T. A.; Holladay, J. E.; White, J. F. *Top Value Added Chemicals from Biomass Volume I — Results of Screening for Potential Candidates from Sugars and Synthesis Gas*; PNNL-14808; BM0102070; TRN: US200807%%599; 2004.

[15] Web of Science [v.7.27.2] - All Databases Citation Report. apps.webofknowledge.com (accessed 2/8/2018).

[16] Banerjee, A.; Dick, G. R.; Yoshino, T.; Kanan, M. W., Carbon dioxide utilization via carbonate-promoted C-H carboxylation. *Nature* **2016,** *531* (7593), 215-9.

[17] Dick, G. R.; Frankhouser, A. D.; Banerjee, A.; Kanan, M. W., A scalable carboxylation route to furan-2,5-dicarboxylic acid. *Green Chem.* **2017,** *19* (13), 2966-2972.

[18] Gevo Ships Renewable Para-Xylene to Toray. Gevo, Inc.: Englewood, Colorodo, 2014.

[19] Gevo Signs Letter of Intent With Highlands EnviroFuels to License Isobutanol Technology for Florida Plant; Plant to Convert Sugar Cane and Sweet Sorghum to 20-25 Million Gallons Per Year of Isobutanol. Gevo, Inc.: Englewood, Colorado, 2014.

[20] DuPont Industrial Biosciences and ADM Announce Breakthrough Platform Technology for Long Sought-After Molecule. E. I. du Pont de Nemours and Company: Wilmington, Deleware, 2016.

[21] Provine, W. D. Cellulosic Biofuels: Today's Opportunity. https://www.acs.org/content/dam/acsorg/policy/acsonthehill/briefings/cellulosicbiofuels/provine-presentation.pdf (accessed February 9, 2018).

[22] American Chemical Society hosts Capitol Hill briefing on Cellulosic Biofuels on Jan. 30. American Chemical Society: Washington, DC, 2012.

[23] Gevo and South Hampton Resources to Build Hydrocarbon Processing Demonstration Plant. Business Wire, Inc.: Englewood, Colorado, 2011.

[24] Mark, C. G.; Fredrik, N.; Bhuma, R. Nucleated Crystallization Of Poly(trimethylene-2,5-furandicarboxylate) (ptf) And Articles Made Therefrom. US 2016/0312001 A1, 2016.

[25] Giardino, C. J.; Griffith, D. B.; Chungfah, H. H.; Howell, J. M.; Watkins, M. H.; Duffy, J. J. Continuous Process For Producing Poly(Trimethylene Terephthalate). US 6353062 B1, 2002.

[26] Gruter, G. J. M.; Dautzenberg, F. Method for the Synthesis of 5-alkoxymethyl Furfural Ethers and Their Use. US 8338626 B2, 2012.

[27] Diego, C. M. D.; Dam, M. A.; Gruter, G. J. M. Method for the Preparation of 2,5-Furandicarboxylic Acid and for the Preparation of the Dialkyl Ester of 2,5-Furandicarboxylic Acid. US 8865921 B2, 2014.

[28] Ruijssenaars, H. J.; Wierckx, N. J. P.; Koopman, F. W.; Straathof, A. J. J.; Winde, J. H. d. Polypeptides Having Oxidoreductase Activity and Their Uses. US 9045787 B2, 2015.

[29] Kania, J.; Blommel, P.; Woods, E.; Dally, B.; Lyman, W.; Cortright, R. Production Of Distillate Fuels From Biomass-derived Polyoxygenates. US 9593054 B2, 2017.

[30] Sanborn, A. J. Processes for the Preparation and Purification of Hydroxymethyl Furaldehyde and Derivatives. WO 2006/063220 A2, 2006.

[31] Sanborn, A. J. Oxidation of Furfural Compounds. WO 2010/132740 A2, 2010.

[32] Diego, C. M. D.; Schammel, W. P.; Dam, M. A.; Gruter, G. J. M. Method for the Preparation of 2,5-Furandicarboxylic Acid and Esters Thereof. WO 2011/043660 A2, 2011.

[33] Jensen, J. S.; Grotkjaer, T.; Pedersen, S.; Ringborg, R. Production of 5-Hydroxymethylfurfural from Fructose Using a Single-Phase Mixed Aqueous-Organic Solvent System. WO 2013/053816 A1, 2013.

[34] Chan, J. W.; Nederberg, F.; Bhuma, R.; Williams, S. R.; Cobb, M. W. Furan Based Polyamides. WO 2013/149180 A1, 2013.

[35] Kalum, L.; Morant, M. D.; Lund, H.; Jensen, J.; Lapainaite, I.; Soerensen, N. H.; Pedersen, S.; Østergaard, L. H.; Xu, F. Enzymatic Oxidation Of 5-hydroxymethylfurfural and Derivatives Thereof. WO 2014/015256 A2, 2014.

[36] Stensrud, K.; Venkitasubramanian, P. Esterification of 2,5-Furan-Dicarboxylic Acid. WO 2014/099438 A2, 2014.

[37] Duursma, A.; Aberson, R.; Smith, D. D.; Flores, J.; DAM, M. A.; Gruter, G. J. M. Process for Preparing a Furan-based Polyamide, a Furan-based Oligomer and Compositions and Articles Comprising the Furan-based Polyamide. WO 2015/060718 A1, 2015.

[38] Ma, J.; Yu, X. F.; Xu, J.; Pang, Y., Synthesis and crystallinity of poly(butylene 2,5-furandicarboxylate). *Polymer* **2012,** *53* (19), 4145-4151.

[39] Gomes, M.; Gandini, A.; Silvestre, A. J. D.; Reis, B., Synthesis and Characterization of Poly(2,5-furan dicarboxylate)s Based on a Variety of Diols. *J. Polym. Sci., Part A: Polym. Chem.* **2011,** *49* (17), 3759-3768.

[40] Ma, J.; Pang, Y.; Wang, M.; Xu, J.; Ma, H.; Nie, X., The copolymerization reactivity of diols with 2,5-furandicarboxylic acid for furan-based copolyester materials. *J. Mater. Chem.* **2012,** *22* (8), 3457-3461.

[41] Jiang, M.; Liu, Q.; Zhang, Q.; Ye, C.; Zhou, G. Y., A series of furan-aromatic polyesters synthesized via direct esterification method based on renewable resources. *J. Polym. Sci., Part A: Polym. Chem.* **2012,** *50* (5), 1026-1036.

[42] Bajpai, P., *Structure of Lignocellulosic Biomass*. Springer Singapore: Singapore, 2016.

[43] Olsson, L.; Hahn-Hagerdal, B., Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme and Microbial Technology* **1996,** *18* (5), 312-331.

[44] Elbehri, A.; Segerstedt, A.; Liu, P., Biofuels and the sustainability challenge: a global assessment of sustainability issues, trends and policies for biofuels and related feedstocks. Trade and Markets Division, Food and Agriculture Organization of the United Nations: Rome, 2013.

[45] HLPE, Biofuels and Food Security. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. Rome, 2013.

[46] Rosatella, A. A.; Simeonov, S. P.; Frade, R. F. M.; Afonso, C. A. M., 5-Hydroxymethylfurfural (HMF) as a building block platform: Biological properties, synthesis and synthetic applications. *Green Chem.* **2011,** *13* (4), 754-793.

[47] Pan, T.; Deng, J.; Xu, Q.; Zuo, Y.; Guo, Q. X.; Fu, Y., Catalytic Conversion of Furfural into a 2,5-Furandicarboxylic Acid-Based Polyester with Total Carbon Utilization. *ChemSusChem* **2013,** *6* (1), 47-50.

[48] Wilson, W. C., 2-Furancarboxylic acid and 2-furylcarbinol. *Org. Synth.* **1926,** *6*, 44-47.

[49] Fischer, R.; Fiserova, M., One-step synthesis of furan-2,5-dicarboxylic acid from furan-2-carboxylic acid using carbon dioxide. *ARKIVOC* **2013**, 405-412.

[50] Wierckx, N.; Koopman, F.; Bandounas, L.; de Winde, J. H.; Ruijssenaars, H. J., Isolation and characterization of Cupriavidus basilensis HMF14 for biological removal of inhibitors from lignocellulosic hydrolysate. *Microb. Biotechnol.* **2010,** *3* (3), 336-43.

[51] Cserháti, M.; Kriszt, B.; Szoboszlay, S.; Tóth, Á.; Szabó, I.; Táncsics, A.; Nagy, I.; Horváth, B.; Nagy, I.; Kukolya, J., De Novo Genome Project of Cupriavidus basilensis OR16. *Journal of Bacteriology* **2012,** *194* (8), 2109-2110.

[52] Koopman, F.; Wierckx, N.; de Winde, J. H.; Ruijssenaars, H. J., Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of Cupriavidus basilensis HMF14. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (11), 4919-24.

[53] Koopman, F. W.; Ruijssenaars, H. J.; Wierckx, N. J. P.; Winde, J. H. d. Novel Microorganism And its Use in Lignocellulose Detoxification. WO 2011/026906 A2, 2011.

[54] Koopman, F.; Wierckx, N.; de Winde, J. H.; Ruijssenaars, H. J., Efficient whole-cell biotransformation of 5-(hydroxymethyl)furfural into FDCA, 2,5-furandicarboxylic acid. *Bioresour. Technol.* **2010,** *101* (16), 6291-6296.

[55] Yong, X.; Chen, R. Method for Preparing Furoic Acid by Cell Catalysis of Furfural. CN 103911405 A, 2014.

[56] Zhou, X.; Zhou, X.; Xu, Y.; Chen, R. R., Gluconobacter oxydans (ATCC 621H) catalyzed oxidation of furfural for detoxification of furfural and bioproduction of furoic acid. *Journal of Chemical Technology & Biotechnology* **2017,** *92* (6), 1285-1289.

[57] Mitsukura, K.; Sato, Y.; Yoshida, T.; Nagasawa, T., Oxidation of heterocyclic and aromatic aldehydes to the corresponding carboxylic acids by Acetobacter and Serratia strains. *Biotechnol. Lett.* **2004,** *26* (21), 1643-1648.

[58] Dijkman, W. P.; Fraaije, M. W., Discovery and Characterization of a 5-Hydroxymethylfurfural Oxidase from Methylovorus sp Strain MP688. *Appl. Environ. Microbiol.* **2014,** *80* (3), 1082-1090.

[59] Dijkman, W. P.; Groothuis, D. E.; Fraaije, M. W., Enzyme-Catalyzed Oxidation of 5-Hydroxymethylfurfural to Furan-2,5-dicarboxylic Acid. *Angew. Chem., Int. Ed.* **2014**.

[60] Ruijssenaars, H. J. Dehydrogenase-Catalysed Production of FDCA. WO 2016/133384 A1, 2016.

[61] Dach, R.; Song, J. J.; Roschangar, F.; Samstag, W.; Senanayake, C. H., The Eight Criteria Defining a Good Chemical Manufacturing Process. *Organic Process Research & Development* **2012,** *16* (11), 1697-1706.

[62] Constable, D. J. C.; Curzons, A. D.; Cunningham, V. L., Metrics to 'green' chemistry—which are the best? *Green Chem.* **2002,** *4* (6), 521-527.

[63] NTP (National Toxicology Program), Report on Carcinogens, Fourteenth Edition. U.S. Department of Health and Human Services: Research Triangle Park, 2016.

[64] Pinkerton, L. E.; Hein, M. J.; Stayner, L. T., Mortality among a cohort of garment workers exposed to formaldehyde: an update. *Occupational and Environmental Medicine* **2004,** *61* (3), 193.

[65] Caspi, R.; Billington, R.; Fulcher, C. A.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Midford, P. E.; Ong, Q.; Ong, W. K.; Paley, S.; Subhraveti, P.; Karp, P. D., The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res* **2018,** *46* (D1), D633-D639.

[66] Caspi, R.; Billington, R.; Ferrer, L.; Foerster, H.; Fulcher, C. A.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; Ong, Q.; Paley, S.; Subhraveti, P.; Weaver, D. S.; Karp, P. D., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **2016,** *44* (D1), D471-80.

[67] Caspi, R.; Altman, T.; Billington, R.; Dreher, K.; Foerster, H.; Fulcher, C. A.; Holland, T. A.; Keseler, I. M.; Kothari, A.; Kubo, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; Ong, Q.; Paley, S.; Subhraveti, P.; Weaver, D. S.; Weerasinghe, D.; Zhang, P.; Karp, P. D., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **2014,** *42* (Database issue), D459-71.

[68] Jankowski, M. D.; Henry, C. S.; Broadbelt, L. J.; Hatzimanikatis, V., Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* **2008,** *95* (3), 1487-1499.

[69] Alberty, R. A., *Thermodynamics of biochemical reactions*. Wiley-Interscience: Hoboken, N.J., 2003.

[70] Cox, J. D.; Wagman, D. D.; Medvedev, V. A., *CODATA key values for thermodynamics*. Hemisphere Pub. Corp.: 1989.

[71] Dijkman, W. P.; Binda, C.; Fraaije, M. W.; Mattevi, A., Structure-Based Enzyme Tailoring of 5-Hydroxymethylfurfural Oxidase. *ACS Catal.* **2015**, 1833-1839.

[72] Glueck, S. M.; Gumus, S.; Fabian, W. M. F.; Faber, K., Biocatalytic carboxylation. *Chem. Soc. Rev.* **2010,** *39* (1), 313-328.

[73] Pesci, L.; Glueck, S. M.; Gurikov, P.; Smirnova, I.; Faber, K.; Liese, A., Biocatalytic carboxylation of phenol derivatives: kinetics and thermodynamics of the biological Kolbe–Schmitt synthesis. *FEBS J.* **2015,** *282* (7), 1334-1345.

[74] Wuensch, C.; Schmidt, N.; Gross, J.; Grischek, B.; Glueck, S. M.; Faber, K., Pushing the equilibrium of regio-complementary carboxylation of phenols and hydroxystyrene derivatives. *J. Biotechnol.* **2013,** *168* (3), 264-70.

[75] Wuensch, C.; Glueck, S. M.; Gross, J.; Koszelewski, D.; Schober, M.; Faber, K., Regioselective Enzymatic Carboxylation of Phenols and Hydroxystyrene Derivatives. *Org. Lett.* **2012,** *14* (8), 1974-1977.

[76] Wuensch, C.; Gross, J.; Steinkellner, G.; Lyskowski, A.; Gruber, K.; Glueck, S. M.; Faber, K., Regioselective ortho-carboxylation of phenols catalyzed by benzoic acid decarboxylases: a biocatalytic equivalent to the Kolbe–Schmitt reaction. *RSC Adv.* **2014,** *4* (19), 9673.

[77] Wieser, M.; Yoshida, T.; Nagasawa, T., Carbon dioxide fixation by reversible pyrrole-2-carboxylate decarboxylase and its application. *J. Mol. Catal. B: Enzym.* **2001,** *11* (4-6), 179-184.

[78] Wieser, M.; Fujii, N.; Yoshida, T.; Nagasawa, T., Carbon dioxide fixation by reversible pyrrole-2-carboxylate decarboxylase from Bacillus megaterium PYR2910. *Eur. J. Biochem.* **1998,** *257* (2), 495-499.

[79] Omura, H.; Wieser, M.; Nagasawa, T., Pyrrole-2-carboxylate decarboxylase from Bacillus megaterium PYR2910, an organic-acid-requiring enzyme. *Eur. J. Biochem.* **1998,** *253* (2), 480-484.

[80] Ulijn, R. V.; Janssen, A. E. M.; Moore, B. D.; Halling, P. J., Predicting when Precipitation-Driven Synthesis Is Feasible: Application to Biocatalysis. *Chem. - Eur. J.* **2001,** *7* (10), 2089-2098.

[81] Berry, D. A.; Ng, K. M., Synthesis of reactive crystallization processes. *Aiche Journal* **1997,** *43* (7), 1737-1750.

[82] Portis, A. R.; Parry, M. A. J., Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth. Res.* **2007,** *94* (1), 121-143.

[83] Miller, E. N.; Jarboe, L. R.; Turner, P. C.; Pharkya, P.; Yomano, L. P.; York, S. W.; Nunn, D.; Shanmugam, K. T.; Ingram, L. O., Furfural Inhibits Growth by Limiting Sulfur Assimilation in Ethanologenic Escherichia coli Strain LY180. *Appl. Environ. Microbiol.* **2009,** *75* (19), 6132-6141.

[84] Trudgill, P. W., Metabolism of 2-Furoic Acid by Pseudomonas F2. *Biochem. J.* **1969,** *113* (4), 577-&.

[85] Kanehisa, M.; Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000,** *28* (1), 27-30.

[86] Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K., KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017,** *45* (D1), D353-D361.

[87] Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M., KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016,** *44* (D1), D457-D462.

[88] Howlett, B. J.; Bar-Tana, J., Polyprenyl p-hydroxybenzoate carboxylase in flagellation of Salmonella typhimurium. *J Bacteriol* **1980,** *143* (2), 644-51.

[89] Leppik, R. A.; Young, I. G.; Gibson, F., Membrane-associated reactions in ubiquinone biosynthesis in Escherichia coli. 3-Octaprenyl-4-hydroxybenzoate carboxy-lyase. *Biochim Biophys Acta* **1976,** *436* (4), 800-10.

[90] Gulmezian, M.; Hyman, K. R.; Marbois, B. N.; Clarke, C. F.; Javor, G. T., The role of UbiX in Escherichia coli coenzyme Q biosynthesis. *Arch Biochem Biophys* **2007,** *467* (2), 144-53.

[91] White, M. D.; Payne, K. A. P.; Fisher, K.; Marshall, S. A.; Parker, D.; Rattray, N. J. W.; Trivedi, D. K.; Goodacre, R.; Rigby, S. E. J.; Scrutton, N. S.; Hay, S.; Leys, D., UbiX is a flavin prenyltransferase required for bacterial ubiquinone biosynthesis. *Nature* **2015,** *522* (7557), 502-506.

[92] Payne, K. A.; White, M. D.; Fisher, K.; Khara, B.; Bailey, S. S.; Parker, D.; Rattray, N. J.; Trivedi, D. K.; Goodacre, R.; Beveridge, R.; Barran, P.; Rigby, S. E.; Scrutton, N. S.; Hay, S.; Leys, D., New cofactor supports alpha,beta-unsaturated acid decarboxylation via 1,3-dipolar cycloaddition. *Nature* **2015,** *522* (7557), 497-501.

[93] Marshall, S. A.; Fisher, K.; Ni Cheallaigh, A.; White, M. D.; Payne, K. A.; Parker, D. A.; Rigby, S. E.; Leys, D., Oxidative Maturation and Structural Characterization of Prenylated FMN Binding by UbiD, a Decarboxylase Involved in Bacterial Ubiquinone Biosynthesis. *J. Biol. Chem.* **2017,** *292* (11), 4623-4637.

[94] Ferguson, K. L.; Arunrattanamook, N.; Marsh, E. N., Mechanism of the Novel Prenylated Flavin-Containing Enzyme Ferulic Acid Decarboxylase Probed by Isotope Effects and Linear Free-Energy Relationships. *Biochemistry-Us* **2016,** *55* (20), 2857-63.

[95] Marshall, S. A.; Payne, K. A. P.; Leys, D., The UbiX-UbiD system: The biosynthesis and use of prenylated flavin (prFMN). *Arch Biochem Biophys* **2017,** *632*, 209-221.

[96] Bailey, S. S.; Payne, K. A. P.; Fisher, K.; Marshall, S. A.; Cliff, M. J.; Spiess, R.; Parker, D. A.; Rigby, S. E.; Leys, D., The role of conserved residues in Fdc decarboxylase in prenylated flavin mononucleotide oxidative maturation, cofactor isomerisation and catalysis. *J. Biol. Chem.* **2017,** *293* (7), 2272-2287.

[97] Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C., SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003,** *31* (13), 3381-3385.

[98] UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017,** *45* (D1), D158-D169.

[99] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-242.

[100] Remmert, M.; Biegert, A.; Hauser, A.; Soding, J., HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **2011,** *9* (2), 173-5.

[101] The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

[102] Malakhov, M. P.; Mattern, M. R.; Malakhova, O. A.; Drinker, M.; Weeks, S. D.; Butt, T. R., SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics* **2004,** *5* (1), 75-86.

[103] Michaelis, L.; Schubert, M. P.; Smythe, C. V., Potentiometric Study of the Flavins. *J. Biol. Chem.* **1936,** *116* (2), 587-607.

[104] Sandler, S. I., *Chemical, biochemical, and engineering thermodynamics*. 4th ed.; John Wiley: Hoboken, N.J., 2006.

[105] McDonald, M. A.; Bommarius, A. S.; Rousseau, R. W., Enzymatic reactive crystallization for improving ampicillin synthesis. *Chemical Engineering Science* **2017,** *165*, 81-88.

[106] Bahn, C. B., Chemical effects on PWR sump strainer blockage after a loss-of-coolant accident: review on U.S. research efforts. *Nuclear Engineering and Technology* **2013,** *45* (3), 295-310.

[107] Apelblat, A.; Manzurola, E., Solubility of Ascorbic, 2-Furancarboxylic, Glutaric, Pimelic, Salicyclic, and O-Phthalic Acids in Water from 279.15 to 342.15 K, and Apparent Molar Volumes of Ascorbic, Glutaric, and Pimelic Acids in Water at 298.15 K. *J. Chem. Thermodyn.* **1989,** *21* (9), 1005-1008.

[108] Vasiliu, M.; Guynn, K.; Dixon, D. A., Prediction of the Thermodynamic Properties of Key Products and Intermediates from Biomass. *J. Phys. Chem. C* **2011,** *115* (31), 15686-15702.

[109] Schroder, B.; Santos, L. M. N. B. F.; Marrucho, I. M.; Coutinho, J. A. P., Prediction of aqueous solubilities of solid carboxylic acids with COSMO-RS. *Fluid Phase Equilib.* **2010,** *289* (2), 140-147.

[110] Institute for Basic Standards.; Wagman, D. D., *Selected values of chemical thermodynamic properties : tables for the first thirty-four elements in the standard order of arrangement*. Dept. of Commerce, National Bureau of Standards : for sale by the Supt. of Docs., U.S. Govt. Print. Off.: Washington, 1968; p iv, 270 p.

[111] Wagman, D. D.; Evans, W. H.; Parker, V. B.; Schumm, R. H.; Halow, I.; Bailey, S. M.; Churney, K. L.; Nuttall, R. L., The NBS Tables of Chemical Thermodynamic Properties - Selected Values for Inorganic and C-1 and C-2 Organic-Substances in Si Units. *J. Phys. Chem. Ref. Data* **1982,** *11*.

[112] Gokel, G. W.; Dean, J. A., *Dean's handbook of organic chemistry*. 2nd ed.; McGraw-Hill: New York, 2004.

[113] Green, J. H. S., Thermodynamic Properties of Organic Oxygen Compounds. *Q. Rev., Chem. Soc.* **1961,** *15* (2), 125-152.

[114] Green, J. H. S.; Harrison, D. J., Spectroscopic and thermodynamic properties of furan derivatives—I. 2- and 2,5-substituted compounds. *Spectrochim. Acta, Part A* **1977,** *33* (9), 843-848.

[115] Roux, M. V.; Temprado, M.; Jiménez, P.; Foces-Foces, C.; García, M. V.; Redondo, M. I., 2- and 3-furancarboxylic acids: a comparative study using calorimetry, IR spectroscopy and X-ray crystallography. *Thermochim. Acta* **2004,** *420* (1-2), 59-66.

[116] Roux, M. V.; Temprado, M.; Jiménez, P.; Pérez-Parajón, J.; Notario, R., Thermochemistry of furancarboxylic acids. *J. Phys. Chem. A* **2003,** *107* (51), 11460-11467.

[117] Goldberg, R. N.; Manley, J. L.; Nuttall, R. L., NBS Technical Note 1206: GAMPHI — a database of activity and osmotic coefficients for aqueous electrolyte solutions. National Bureau of Standards, U.S. Department of Commerce: Gaithersburgh, MD, 1985.

[118] Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G., Minnesota Solvation Database – version 2012. University of Minnesota,: Minneapolis, MN, 2012.

[119] Onken, U.; Rarey-Nies, J.; Gmehling, J., The Dortmund Data Bank: A computerized system for retrieval, correlation, and prediction of thermodynamic properties of mixtures. *International Journal of Thermophysics* **1989,** *10* (3), 739-747.

[120] Pitzer, K. S., Thermodynamics of Electrolytes. I. Theoretical Basis and General Equations. *Journal of Physical Chemistry* **1973,** *77* (2), 268-277.

[121] Pitzer, K. S.; Mayorga, G., Thermodynamics of Electrolytes. II. Activity and Osmotic Coefficients for Strong Electrolytes with One or Both Ions Univalent. *Journal of Physical Chemistry* **1973,** *77* (19), 2300-2308.

[122] Pitzer, K. S.; Mayorga, G., Thermodynamics of Electrolytes. III. Activity and Osmotic Coefficients for 2-2 Electrolytes. *Journal of Solution Chemistry* **1974,** *3* (7), 539-546.

[123] Pitzer, K. S.; Kim, J. J., Thermodynamics of Electrolytes. IV. Activity and Osmotic Coefficients for Mixed Electrolytes. *J. Am. Chem. Soc.* **1974,** *96* (18), 5701-5707.

[124] Pitzer, K. S., Thermodynamics of Electrolytes. V. Effects of Higher-Order Electrostatic Terms. *Journal of Solution Chemistry* **1975,** *4* (3), 249-265.

[125] Pitzer, K. S.; Silvester, L. F., Thermodynamics of Electrolytes. VI. Weak Electrolytes Including H3PO4. *Journal of Solution Chemistry* **1976,** *5* (4), 269-278.

[126] Pitzer, K. S.; Roy, R. N.; Silvester, L. F., Thermodynamics of Electrolytes. 7. Sulfuric-Acid. *J. Am. Chem. Soc.* **1977,** *99* (15), 4930-4936.

[127] Silvester, L. F.; Pitzer, K. S., Thermodynamics of Electrolytes. 8. High-Temperature Properties, Including Enthalpy and Heat-Capacity, with Application to Sodium-Chloride. *Journal of Physical Chemistry* **1977,** *81* (19), 1822-1828.

[128] Pitzer, K. S.; Peterson, J. R.; Silvester, L. F., Thermodynamics of Electrolytes. IX. Rare-Earth Chlorides, Nitrates, and Perchlorates. *Journal of Solution Chemistry* **1978,** *7* (1), 45-56.

[129] Silvester, L. F.; Pitzer, K. S., Thermodynamics of Electrolytes. X. Enthalpy and Effect of Temperature on Activity-Coefficients. *Journal of Solution Chemistry* **1978,** *7* (5), 327-337.

[130] Pitzer, K. S.; Silvester, L. F., Thermodynamics of Electrolytes. 11. Properties of 3-2, 4-2, and Other High-Valence Types. *Journal of Physical Chemistry* **1978,** *82* (11), 1239-1242.

[131] Bradley, D. J.; Pitzer, K. S., Thermodynamics of Electrolytes. 12. Dielectric Properties of Water and Debye-Huckel Parameters to 350-Degrees-C and 1-Kbar. *Journal of Physical Chemistry* **1979,** *83* (12), 1599-1603.

[132] Haghtalab, A.; Peyvandi, K., Generalized Electrolyte-UNIQUAC-NRF model for calculation of solubility and vapor pressure of multicomponent electrolytes solutions. *Journal of Molecular Liquids* **2012,** *165*, 101-112.

[133] Clegg, S. L.; Seinfeld, J. H., Thermodynamic models of aqueous solutions containing inorganic electrolytes and dicarboxylic acids at 298.15 K. 1. The acids as nondissociating components. *J Phys Chem A* **2006,** *110* (17), 5692-717.

[134] Clegg, S. L.; Seinfeld, J. H., Thermodynamic models of aqueous solutions containing inorganic electrolytes and dicarboxylic acids at 298.15 K. 2. Systems including dissociation equilibria. *J Phys Chem A* **2006,** *110* (17), 5718-34.

[135] Gross, J.; Sadowski, G., Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains. *Fluid Phase Equilib.* **2000,** *168* (2), 183-199.

[136] Gross, J.; Sadowski, G., Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.* **2001,** *40* (4), 1244-1260.

[137] Held, C.; Sadowski, G., Thermodynamics of Bioreactions. *Annu. Rev. Chem. Biomol. Eng.* **2016,** *7*, 395-414.

[138] Rothmund, V., Die gegenseitige Löslichkeit von Flüssigkeiten und der kritische Lösungspunkt. *Z. Phys. Chem.* **1898,** *3*, 433-492.

[139] Mains, G. H., The System Furfural-Water - I. A study of its properties with reference to their commercial application in the production of furfural. *Chem. Metall. Eng.* **1922,** *26* (17), 779-784.

[140] Evans, W. V.; Aylesworth, M. B., Some critical constants of furfural. *Ind. Eng. Chem.* **1926,** *18*, 24-27.

[141] Landrieiu, P.; Baylocq, F.; Johnson, J. R., Etude thermochimique dans la serie furanique. *Bull. Soc. Chim. Fr.* **1929,** *45*, 36-49.

[142] Cativiela, C.; Dejardin, J. L.; Elguero, J.; Garcia, J. I.; Gonzalez, E.; Mayoral, J. A., Acidity in Water (Pka Values) of Carboxylic-Acids Derived from Simple Heterocycles (Azoles and Azines). *Collect. Czech. Chem. Commun.* **1990,** *55* (1), 72-79.

[143] Guo, Y.; Yin, Q.; Hao, H.; Zhang, M.; Bao, Y.; Hou, B.; Chen, W.; Zhang, H.; Cong, W., Measurement and Correlation of Solubility and Dissolution Thermodynamic Properties of Furan-2-carboxylic Acid in Pure and Binary Solvents. *J. Chem. Eng. Data* **2014,** *59* (4), 1326-1333.

[144] Lilga, M. A.; Hallen, R. T.; Gray, M., Production of Oxidized Derivatives of 5-Hydroxymethylfurfural (HMF). *Top. Catal.* **2010,** *53* (15-18), 1264-1269.

[145] Kuramochi, H.; Noritomi, H.; Hoshino, D.; Nagahama, K., Measurements of Solubilities of Two Amino Acids in Water and Prediction by the UNIFAC Model. *Biotechnol. Progr.* **1996,** *12* (3), 371-379.

[146] Kurosawa, I.; Teja, A. S.; Rousseau, R. W., Solid-liquid equilibria in L-leucine plus L-valine plus water. *Fluid Phase Equilib.* **2004,** *224* (2), 245-249.

[147] Soto, A.; Arce, A.; Khoshkbarchi, M. K.; Vera, J. H., Measurements and modelling of the solubility of a mixture of two amino acids in aqueous solutions. *Fluid Phase Equilib.* **1999,** *160*, 893-901.

[148] Ji, P. J.; Feng, W., Solubility of amino acids in water and aqueous solutions by the statistical associating fluid theory. *Ind. Eng. Chem. Res.* **2008,** *47* (16), 6275-6279.

[149] Gross, J.; Sadowski, G., Application of the Perturbed-Chain SAFT Equation of State to Associating Systems. *Ind. Eng. Chem. Res.* **2002,** *41* (22), 5510-5515.

[150] Held, C.; Sadowski, G.; Carneiro, A.; Rodríguez, O.; Macedo, E. A., Modeling thermodynamic properties of aqueous single-solute and multi-solute sugar solutions with PC-SAFT. *AIChE Journal* **2013,** *59* (12), 4794-4805.

[151] Daldrup, J.-B. G.; Held, C.; Sadowski, G.; Schembecker, G., Modeling pH and Solubilities in Aqueous Multisolute Amino Acid Solutions. *Ind. Eng. Chem. Res.* **2011,** *50* (6), 3503-3509.

[152] Prausnitz, J. M., *Molecular thermodynamics of fluid-phase equilibria*. Prentice-Hall: Englewood Cliffs, N.J., 1969; p xiii, 523.

[153] Gmehling, J. G.; Anderson, T. F.; Prausnitz, J. M., Solid-Liquid Equilibria Using Unifac. *Ind. Eng. Chem. Fundam.* **1978,** *17*, 269-273.

[154] Wolbach, J. P.; Sandler, S. I., Using molecular orbital calculations to describe the phase behavior of cross-associating mixtures. *Ind. Eng. Chem. Res.* **1998,** *37* (8), 2917-2928.

[155] Fuchs, D.; Fischer, J.; Tumakaka, F.; Sadowski, G., Solubility of amino acids: Influence of the pH value and the addition of alcoholic cosolvents on aqueous solubility. *Ind. Eng. Chem. Res.* **2006,** *45* (19), 6578-6584.

[156] Sun, M. S.; Harriss, D. K.; Magnuson, V. R., Activity Corrections for Ionic Equilibria in Aqueous Solutions. *Can. J. Chem.* **1980,** *58* (12), 1253-1257.

[157] Carneiro, A. P.; Held, C.; Rodriguez, O.; Sadowski, G.; Macedo, E. A., Solubility of Sugars and Sugar Alcohols in Ionic Liquids: Measurement and PC-SAFT Modeling. *J. Phys. Chem. B.* **2013,** *117* (34), 9980-9995.

[158] Liessmann, G.; Schmidt, W.; Reiffarth, S., Data compilation of the Saechsische Olefinwerke Boehlen Germany. *Recommended Thermophysical Data* **1995**.

[159] Daubert, T. E.; Danner, R. P., *Data Compilation Tables of Properties of Pure Compounds*. Design Institute for Physical Property Data: New York, NY, 1985.

[160] Cameretti, L. F.; Sadowski, G., Modeling of aqueous amino acid and polypeptide solutions with PC-Saft. *Chem. Eng. Process.* **2008,** *47*, 1018-1025.

[161] Tumakaka, F.; Sadowski, G., Application of the Perturbed-Chain SAFT equation of state to polar systems. *Fluid Phase Equilib.* **2004,** *217* (2), 233-239.

[162] Gilman, H.; Burtner, R. R.; Vanderwal, R. J., Orientation in the furan nucleus. IV. The equivalence of the alpha positions and of the beta positions in furan. *Recl. Trav. Chim. Pays-Bas* **1933,** *52*, 151-155.

[163] Timmermanns, J., Investigation of the Freezing Point of Organic Substances VII. *Bull. Soc. Chim. Belg.* **1922,** *31* (389).

[164] Domalski, E. S.; Hearing, E. D., Heat Capacities and Entropies of Organic Compounds in the Condensed Phase. Volume III. *J. Phys. Chem. Ref. Data* **1996,** *25* (1), 1-525.

[165] Caßens, J. Modellierung thermodynamischer Eigenschaften pharmazeutischer Substanzen in Lösungsmitteln und Lösungsmittelgemischen. Dissertation, Universität Dortmund, München, 2013.

[166] Gmehling, J.; Onken, U.; Arlt, W.; Grenzheuser, P.; Kolbe, B.; Weidlich, U.; Rarey-Nies, R. J., *Vapor-Liquid-Equilibrium Data Collection, Vol. I*. DECHEMA Chemistry Data Series: Frankfurt 1977-1988.

[167] Gilbert, A. S., Entropy–enthalpy compensation in the fusion of organic molecules: implications for Walden's rule and molecular freedom in the liquid state. *Thermochim. Acta* **1999,** *339* (1-2), 131-142.

[168] Ruether, F.; Sadowski, G., Modeling the solubility of pharmaceuticals in pure solvents and solvent mixtures for drug process design. *J. Pharm. Sci.* **2009,** *98*, 4205-4255.

[169] Holmquist, M., Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci.* **2000,** *1* (2), 209-235.

[170] Blum, J. K.; Bommarius, A. S., Amino ester hydrolase from *Xanthomonas campestris* pv. campestris, ATCC 33913 for enzymatic synthesis of ampicillin. *Journal of Molecular Catalysis B-Enzymatic* **2010,** *67* (1-2), 21-28.

[171] Gao, D.; Narasimhan, D. L.; Macdonald, J.; Brim, R.; Ko, M.-C.; Landry, D. W.; Woods, J. H.; Sunahara, R. K.; Zhan, C.-G., Thermostable variants of cocaine esterase for long-time protection against cocaine toxicity. *Molecular Pharmacology* **2009,** *75* (2), 318-323.

[172] Rauwerdink, A.; Kazlauskas, R. J., How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of α/β-hydrolase fold enzymes. *ACS Catalysis* **2015,** *5* (10), 6153-6176.

[173] Huang, X.; Gao, D.; Zhan, C. G., Computational design of a thermostable mutant of cocaine esterase via molecular dynamics simulations. *Organic & Biomolecular Chemistry* **2011,** *9* (11), 4138-4143.

[174] Brim, R. L.; Noon, K. R.; Collins, G. T.; Stein, A.; Nichols, J.; Narasimhan, D.; Ko, M. C.; Woods, J. H.; Sunahara, R. K., The fate of bacterial cocaine esterase (CocE): an in vivo study of CocE-mediated cocaine hydrolysis, CocE pharmacokinetics, and CocE elimination. *J. Pharmacol. Exp. Ther.* **2012,** *340* (1), 83-95.

[175] Paye, M. F., Biocatalysis of amide and peptide bond synthesis by cocaine esterase and alpha-amino acid ester hydrolase. *Georgia Institute of Technology* **2017**, 1-149.

[176] Blum, J. K., Broadening the enyzme-catalyzed synthesis of semi-synthetic antibiotics. *Georgia Institute of Technology* **2011**, 1-197.

[177] Janes, L. E.; Löwendahl, A. C.; Kazlauskas, R. J., Quantitative screening of hydrolase libraries using pH indicators: Identifying active and enantioselective hydrolases. *Chemistry – A European Journal* **1998,** *4* (11), 2324-2331.

[178] Yi, D.; Devamani, T.; Abdoul-Zabar, J.; Charmantray, F.; Helaine, V.; Hecquet, L.; Fessner, W.-D., A pH-based high-throughput assay for transketolase: fingerprinting of substrate tolerance and quantitative kinetics. *ChemBioChem* **2012,** *13* (15), 2290-2300.

[179] Deng, C.; Chen, R. R., A pH-sensitive assay for galactosyltransferase. *Analytical Biochemistry* **2004,** *330* (2), 219-226.

[180] Xiao, Y. Z.; Huo, X. D.; Qian, Y.; Zhang, Y.; Chen, G. Q.; Ouyang, P. K.; Lin, Z. L., Engineering of a CPC acylase using a facile pH indicator assay. *Journal of Industrial Microbiology & Biotechnology* **2014,** *41* (11), 1617-1625.

[181] Chapman, E.; Wong, C.-H., A pH sensitive colorometric assay for the high-Throughput screening of enzyme inhibitors and substrates: A case study using kinases. *Bioorganic & Medicinal Chemistry* **2002,** *10* (3), 551-555.

[182] El-Obeid, H. A.; Gad-Kariem, E. A.; Al-Rashood, K. A.; Al-Khamees, H. A.; El-Shafie, F. S.; Bawazeer, G. A. M., A selective colorimetric method for the determination of penicillins and cephalosporins with α-aminoacyl functions. *Analytical Letters* **1999,** *32* (14), 2809-2823.

[183] Pereira, P. C.; Arends, I. W. C. E.; Sheldon, R. A., A green and expedient synthesis of enantiopure diketopiperazines via enzymatic resolution of unnatural amino acids. *Tetrahedron Letters* **2014,** *55* (36), 4991-4993.

[184] Cashman, J. R.; Berkman, C. E.; Underiner, G.; Kolly, C. A.; Hunter, A. D., Cocaine Benzoyl Thioester: Synthesis, Kinetics of Base Hydrolysis, and Application to the Assay of Cocaine Esterases. *Chemical Research in Toxicology* **1998,** *11* (8), 895-901.

[185] Blum, J. K.; Ricketts, M. D.; Bommarius, A. S., Improved thermostability of AEH by combining B-FIT analysis and structure-guided consensus method. *Journal of Biotechnology* **2012,** *160* (3–4), 214-221.

[186] Ellis, K. J.; Morrison, J. F., Buffers of constant ionic strength for studying pH-dependent processes. *Methods Enzymol.* **1982,** *87*, 405-26.

[187] Gans, P.; O'Sullivan, B., GLEE, a new computer program for glass electrode calibration. *Talanta* **2000,** *51* (1), 33-7.

[188] Burgot, J.-L., *Ionic equilibria in analytical chemistry*. Springer Science+Business Media: New York, 2012.

[189] Sawilowski, S. S., Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when. *J. Mod. Appl. Stat Methods* **2002,** *1* (2), 461-472.

[190] Deffeyes, K. S., Carbonate equilibria: a graphic and algebraic approach. *Limnol. Oceanogr.* **1956,** *10* (3), 412-426.

[191] He, F.; Hogan, S.; Latypov, R. F.; Narhi, L. O.; Razinkov, V. I., High throughput thermostability screening of monoclonal antibody formulations. *J Pharm Sci* **2010,** *99* (4), 1707-20.

[192] Dobrogowska, C.; Hepler, L. G.; Apelblat, A., The Enthalpies of Dilution of Aqueous Organic-Acids - Oxalic-Acid and Citric-Acid at 298.15-K. *J. Chem. Thermodyn.* **1990,** *22* (2), 167-172.

[193] Messikomer, E. E.; Wood, R. H., Enthalpy of Dilution of Aqueous Sodium-Chloride at 298.15 to 373.15 K, Measured with a Flow Calorimeter. *J. Chem. Thermodyn.* **1975,** *7* (2), 119-130.

[194] Smith, L. E. S.; Cone, J. T.; Van Hook, W. A., Enthalpies of dilution of aqueous electrolytes: The NaCl/H2O system. *Journal of Solution Chemistry* **1980,** *9* (2), 81-95.

[195] Young, T. F.; Vogel, O. G., The relative heat contents of the constituents of aqueous sodium chloride solutions. *J. Am. Chem. Soc.* **1932,** *54*, 3030-3040.

[196] Millero, F. J.; Duer, W. C.; Shepard, E.; Chetirkin, P. V., Enthalpies of Dilution of Phosphate Solutions at 30-Degrees-C. *Journal of Solution Chemistry* **1978,** *7* (12), 877-889.

[197] Hedwig, G. R.; Watson, I. D., A microcalorimetric study of the association of hexacyanoferrate(II) ions with calcium and magnesium ions in aqueous solution. *Journal of Solution Chemistry* **1981,** *10* (6), 411-418.

[198] Weber, C. F.; Beahm, E. C.; Watson, J. S., Modeling Thermodynamics and Phase Equilibria for Aqueous Solutions of Trisodium Phosphate. *Journal of Solution Chemistry* **1999,** *28* (11), 1207-1238.

[199] Holmes, H. F.; Busey, R. H.; Simonson, J. M.; Mesmer, R. E.; Archer, D. G.; Wood, R. H., The enthalpy of dilution of HCl(aq) to 648 K and 40 MPa thermodynamic properties. *J. Chem. Thermodyn.* **1987,** *19* (8), 863-890.

[200] Bernard, M. A.; Burgot, J. L., Thermometric Titrations of Polyprotic Acids. *Anal. Chem.* **1979,** *51* (13), 2122-2126.

[201] Rowland, D.; May, P. M., Thermodynamics of Strong Aqueous Electrolyte Solutions at t = 25 °C Described by the Hückel Equations. *J. Chem. Eng. Data* **2014,** *59* (6), 2030-2039.

[202] de Rivera, M. R.; Socorro, F.; Matos, J. S., Heats of mixing using an isothermal titration calorimeter: associated thermal effects. *Int J Mol Sci* **2009,** *10* (7), 2911-20.

[203] Ladbury, J. E., Application of isothermal titration calorimetry in the biological sciences: things are heating up! *Biotechniques* **2004,** *37* (6), 885-7.

[204] O'Brien, R.; Ladbury, J. E.; Chowdhry, B. Z., Isothermal Titration Calorimetry of Biomolecules. In *Protein-Ligand Interactions*, Chowdhry, B. Z.; Harding, S. E., Eds. Oxford University Press: Oxford ; New York, 2001; Vol. 1, pp 263-286.

[205] Tajc, S. G.; Tolbert, B. S.; Basavappa, R.; Miller, B. L., Direct determination of thiol pKa by isothermal titration microcalorimetry. *J. Am. Chem. Soc.* **2004,** *126* (34), 10508-9.

[206] Shoghi, E.; Romero, L.; Reta, M.; Rafols, C.; Bosch, E., Enthalpies and constants of dissociation of several neutral and cationic acids in aqueous and methanol/water solutions at various temperatures. *J Pharm Biomed Anal* **2009,** *49* (4), 923-30.

[207] Zheng, Y.; Liu, M.; Wang, C. M.; Xie, J. M.; Sun, D. Z.; Wang, B. Q.; Han, J.; Niu, M. J., Simultaneous determination of enthalpies of dilution and dissociation of protocatechuic acid in aqueous salt solutions by calorimetric measurements. *J. Chem. Thermodyn.* **2015,** *91*, 177-185.

[208] Salgado, L. E. V.; Vargas-Hernández, C., Spectrophotometric Determination of the pKa, Isosbestic Point and Equation of Absorbance vs. pH for a Universal pH Indicator. *American Journal of Analytical Chemistry* **2014,** *05* (17), 1290-1301.

[209] Po, H. N.; Senozan, N. M., The Henderson-Hasselbalch equation: Its history and limitations. *Journal of Chemical Education* **2001,** *78* (11), 1499-1503.

[210] Luff, R.; Haeckel, M.; Wallmann, K., Robust and fast FORTRAN and MATLAB (R) libraries to calculate pH distributions in marine systems. *Computers & Geosciences* **2001,** *27* (2), 157-169.

[211] Smith, R. M.; Martell, A. E.; Motekaitis, R. J., NIST Critically Selected Stability Constants of Metal Complexes Database. In *NIST*, 8 ed.; 2004.

[212] Apelblat, A., *Citric Acid*. Springer International Publishing: Switzerland, 2014.

[213] Christensen, J. J.; Izatt, R. M.; Hansen, L. D., Thermodynamics of Proton Ionization in Dilute Aqueous Solution. VII. [UNK]H° and [UNK]S° Values for Proton Ionization from Carboxylic Acids at 25°. *J. Am. Chem. Soc.* **1967,** *89* (2), 213-222.

[214] McBryde, W. A. E., The pH meter as a hydrogen-ion concentration probe. *Analyst* **1969,** *94* (1118), 337-346.

[215] Reijenga, J.; van Hoof, A.; van Loon, A.; Teunissen, B., Development of Methods for the Determination of pKa Values. *Anal. Chem. Insights* **2013,** *8*, 53-71.

[216] Teixeira, J., *Molecular physical chemistry : a computer-based approach using Mathematica and Gaussian*. Cham, Switzerland : Springer: Cham, Switzerland, 2017.

[217] Eigen, M., Methods for Investigation of Ionic Reactions in Aqueous Solutions with Half-Times as Short as 10-9 Sec - Application to Neutralization and Hydrolysis Reactions. *Discussions of the Faraday Society* **1954,** (17), 194-205.

[218] Eigen, M., Immeasurably fast reactions, Nobel Lecture, December 11 , 1967. In *Nobel Lectures, Chemistry 1963-1970*, Elsevier Publishing Company: Amsterdam, 1972.

[219] Hazen, R., Chance, Necessity, and the Origins of Life. School of Chemical & Biomolecular Engineering Fall Semester 2017 Seminar Series: Georgia Institute of Technology, Atlanta, Georgia, 2017.

[220] Miller, S. L., A Production of Amino Acids under Possible Primitive Earth Conditions. *Science* **1953,** *117* (3046), 528-529.

[221] Hazen, R. M., Chance, necessity and the origins of life: a physical sciences perspective. *Philos Trans A Math Phys Eng Sci* **2017,** *375* (2109).

[222] Haub, C. How many people have ever lived on earth? http://www.prb.org/Publications/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx (accessed February 17, 2017).

[223] Whitman, W. B.; Coleman, D. C.; Wiebe, W. J., Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95* (12), 6578-6583.

[224] Drake, J. W., A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences* **1991,** *88* (16), 7160.

[225] Ouzounis, C. A.; Karp, P. D., Global Properties of the Metabolic Map of Escherichia coli. *Genome Research* **2000,** *10* (4), 568-576.

[226] Sørensen, S. P. L., Enzyme studies - Note II - The measurement and the significance of hydrogenic concentrate in enzymatic processes. *Biochemische Zeitschrift* **1909,** *21*, 131-304.

[227] Utzerath, C.; Fernández, G., Shaping Science for Increasing Interdependence and Specialization. *Trends in Neurosciences 40* (3), 121-124.

[228] Komor, R. S.; Romero, P. A.; Xie, C. B.; Arnold, F. H., Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng., Des. Sel.* **2012,** *25* (12), 827-833.

[229] Steipe, B.; Schiller, B.; Pluckthun, A.; Steinbacher, S., Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. *Journal of Molecular Biology* **1994,** *240* (3), 188-192.

[230] Bourgeois, D.; Weik, M., Kinetic protein crystallography: a tool to watch proteins in action. *Crystallography Reviews* **2009,** *15* (2), 87-118.

[231] Fischer, N.; Neumann, P.; Bock, L.; Ficner, R.; Rodnina, M.; Stark, H., Breaking the 3 angstrom resolution barrier in single particle cryo-EM. *European Biophysics Journal with Biophysics Letters* **2015,** *44*, S46-S46.

[232] Matthies, D.; Dalmas, O.; Borgnia, M. J.; Dominik, P. K.; Merk, A.; Rao, P.; Reddy, B. G.; Islam, S.; Bartesaghi, A.; Perozo, E.; Subramaniam, S., Cryo-EM Structures of the Magnesium Channel CorA Reveal Symmetry Break upon Gating. *Cell* **2016,** *164* (4), 747-756.

[233] Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S., Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **2016,** *165* (7), 1698-1707.

[234] Frank, J., Time-resolved cryo-electron microscopy: Recent progress. *Journal of Structural Biology* **2017,** *200* (3), 303-306.

[235] Northrop, D. B., Effects of high pressure on enzymatic activity. *Biochim Biophys Acta* **2002,** *1595* (1-2), 71-9.

[236] Athès, V.; Combes, D., Effect of high hydrostatic pressure on enzyme stability. In *Progress in Biotechnology*, Ballesteros, A.; Plou, F. J.; Iborra, J. L.; Halling, P. J., Eds. Elsevier: 1998; Vol. 15, pp 205-210.

[237] Karim, A. S.; Dudley, Q. M.; Jewett, M. C., Cell-Free Synthetic Systems for Metabolic Engineering and Biosynthetic Pathway Prototyping. In *Industrial Biotechnology*, Wiley-VCH Verlag GmbH & Co. KGaA: 2017; pp 125-148.

[238] Lopez-Gallego, F.; Schmidt-Dannert, C., Multi-enzymatic synthesis. *Curr Opin Chem Biol* **2010,** *14* (2), 174-83.

[239] Veibel, S., On the Application of P-Carboxyphenylhydrazones in the Identification of Carbonyl Compounds. *Acta Chem. Scand.* **1947,** *1* (1), 54-68.

[240] Grammaticakis, P., Contribution à l'étude de l'absorption dans l'ultraviolet moyen et le visible des aryl- et aroyl- hydrazones. II. Carboxyphénylhydrazones (Ortho, Meta et Para). *Bull. Soc. Chim. Fr.* **1954,** *21*, 1381-1390.

[241] Lococo, F.; Ceccon, L.; Valentini, C.; Novelli, V., High-Performance Liquid-Chromatographic Determination of 2-Furaldehyde in Spirits. *Journal of Chromatography* **1992,** *590* (2), 235-240.

[242] LoCoco, F.; Novelli, V.; Valentini, C.; Ceccon, L., High-performance liquid chromatographic determination of 2-furaldehyde and 5-hydroxymethyl-2-furaldehyde in fruit juices. *Journal of Chromatographic Science* **1997,** *35* (12), 578-583.

[243] LoCoco, F.; Valentini, C.; Novelli, V.; Ceccon, L., High-performance liquid chromatographic determination of 2-furaldehyde and 5-hydroxymethyl-2-furaldehyde in honey. *Journal of Chromatography A* **1996,** *749* (1-2), 95-102.

[244] Lococo, F.; Valentini, C.; Novelli, V.; Ceccon, L., High-Performance Liquid-Chromatography Determination of 2-Furaldehyde and 5-Hydroxymethyl-2-Furaldehyde in Processed Citrus Juices. *Journal of Liquid Chromatography* **1994,** *17* (3), 603-617.

[245] Zhou, M.; Diwu, Z.; Panchuk-Voloshina, N.; Haugland, R. P., A stable nonfluorescent derivative of resorufin for the fluorometric determination of trace hydrogen peroxide: applications in detecting the activity of phagocyte NADPH oxidase and other oxidases. *Anal Biochem* **1997,** *253* (2), 162-8.

[246] Zeitsch, K. J., *The chemistry and technology of furfural and its many by-products*. Elsevier: Amsterdam ; New York, 2000; p xv, 358 p.

[247] Fiedler, J. D.; Brown, S. D.; Lau, J. L.; Finn, M. G., RNA-directed packaging of enzymes within virus-like particles. *Angew Chem Int Ed Engl* **2010,** *49* (50), 9648-51.

[248] Berg, R. L.; Vanderzee, C. E., Thermodynamics of carbon dioxide and carbonic acid: (a) the standard enthalpies of solution of $Na_2CO_{3(s)}$, $NaHCO_{3(s)}$, and $CO_{2(g)}$

in water at 298.15 K; (b) the standard enthalpies of formation, standard Gibbs energies of formation, and standard entropies of $CO_{2(aq)}$, $HCO_3^{-(aq)}$, CO_3^{2-}(aq), $NaHCO_{3(s)}$, $Na_2CO_{3(s)}$, $Na_2CO_3 \cdot H_2O_{(s)}$, and $Na_2CO_3 \cdot 10H_2O_{(s)}$. *J. Chem. Thermodyn.* **1978,** *10* (12), 1113-1136.

[249] Nojavan, S.; Ghassempour, A.; Bashour, Y.; Darbandi, M. K.; Ahmadi, S. H., Determination of residual solvents and investigation of their effect on ampicillin trihydrate crystal structure. *J Pharm Biomed Anal* **2005,** *36* (5), 983-8.

[250] Held, C.; Cameretti, L. F.; Sadowski, G., Modeling aqueous electrolyte solutions: Part 1. Fully dissociated electrolytes. *Fluid Phase Equilib.* **2008,** *270* (1–2), 87-96.

[251] Held, C.; Reschke, T.; Mohammad, S.; Luza, A.; Sadowski, G., ePC-SAFT revised. *Chem. Eng. Res. Des.* **2014,** *92* (12), 2884-2897.

[252] Held, C.; Sadowski, G., Modeling aqueous electrolyte solutions. Part 2. Weak electrolytes. *Fluid Phase Equilib.* **2009,** *279* (2), 141-148.

[253] Lange, L.; Lehmkemper, K.; Sadowski, G., Predicting the Aqueous Solubility of Pharmaceutical Cocrystals As a Function of pH and Temperature. *Crystal Growth & Design* **2016,** *16* (5), 2726-2740.

[254] Lozano, P.; De Diego, T.; Vaultier, M.; Iborra, J. L., Enzyme Catalysis in Ionic Liquids and Supercritical Carbon Dioxide. **2010,** *1038*, 181-196.

[255] Matsuda, T.; Watanabe, K.; Harada, T.; Nakamura, K., Enzymatic reactions in supercritical CO_2: carboxylation, asymmetric reduction and esterification. *Catal. Today* **2004,** *96* (3), 103-111.

[256] Payne, S. M.; Kerton, F. M., Solubility of bio-sourced feedstocks in 'green' solvents. *Green Chem.* **2010,** *12* (9), 1648-1653.

[257] Wimmer, Z.; Zarevucka, M., A review on the effects of supercritical carbon dioxide on enzyme activity. *Int. J. Mol. Sci.* **2010,** *11* (1), 233-53.

[258] Karimi, A.; Golbabaei, F.; Mehrnia, M. R.; Neghab, M.; Mohammad, K.; Nikpey, A.; Pourmand, M. R., Oxygen mass transfer in a stirred tank bioreactor using different impeller configurations for environmental purposes. *Iranian Journal of Environmental Health Science & Engineering* **2013,** *10* (1), 6-6.

[259] Garcia-Ochoa, F.; Gomez, E., Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. *Biotechnol Adv* **2009,** *27* (2), 153-76.

[260] Darunte, L. A.; Walton, K. S.; Sholl, D. S.; Jones, C. W., CO_2 capture via adsorption in amine-functionalized sorbents. *Current Opinion in Chemical Engineering* **2016,** *12,* 82-90.

[261] Sinha, A.; Darunte, L. A.; Jones, C. W.; Realff, M. J.; Kawajiri, Y., Systems Design and Economic Analysis of Direct Air Capture of CO_2 through Temperature Vacuum Swing Adsorption Using MIL-101(Cr)-PEI-800 and mmen-Mg_2(dobpdc) MOF Adsorbents. *Ind. Eng. Chem. Res.* **2017,** *56* (3), 750-764.

[262] Harte, C. R.; Baker, E. M.; Purcell, H. H., Absorption of Carbon Dioxide in Sodium Carbonate-Bicarbonate Solutions. *Industrial & Engineering Chemistry* **1933,** *25* (5), 528-531.

[263] Harte, C. R.; Baker, E. M., Absorption of Carbon Dioxide in Aqueous Sodium Carbonate-Bicarbonate Solutions. *Industrial & Engineering Chemistry* **1933,** *25* (10), 1128-1132.

[264] Harris, J. F. S., J. M. *Engineering Calculations for the Distillation of the Furfural-Water System*; 2070; Forest Products Laboratory, Forest Service U.S. Department of Agriculture: Madison, Wis., 1959.

[265] Curtis, R. G. H., H. H., Equilibria in Furfural-Water Systems under Increased Pressure and the Influence of Added Salts upon the Mutual Solubilities of Furfural and Water. *Aust. J. Sci. Res., Ser. A* **1948,** *vol. 1*.

[266] Nhien, L. C.; Long, N. V. D.; Kim, S.; Lee, M., Techno-economic assessment of hybrid extraction and distillation processes for furfural production from lignocellulosic biomass. *Biotechnology for Biofuels* **2017,** *10* (1), 81.

[267] Santana, M.; Ribeiro, M. P. A.; Leite, G. A.; Giordano, R. L. C.; Giordano, R. C.; Mattedi, S., Solid–liquid equilibrium of substrates and products of the enzymatic synthesis of ampicillin. *AIChE Journal* **2010,** *56* (6), 1578-1583.

[268] Das Gupta, V.; Reed, J. B., Jr., First pKa values of some acid-base indicators. *J. Pharm. Sci.* **1970,** *59* (11), 1683-5.

[269] Andon, R. J. L.; Cox, J. D.; Herington, E. F. G., Phase relationships in the pyridine series. Part V. The thermodynamic properties of dilute solutions of pyridine bases in water at 25° and 40°. *Journal of the Chemical Society (Resumed)* **1954,** (0), 3188-3196.

[270] Scriven, E. F.; Murugan, R., Pyridine and pyridine derivatives. *Kirk-Othmer Encyclopedia of Chemical Technology* **2005**.

[271] Pure, I. U. o.; Data, A. C. C. o. E.; Serjeant, E. P.; Dempsey, B.; Pure, I. U. o.; Data, A. C. C. o. E., *Ionisation constants of organic acids in aqueous solution*. Pergamon Press: 1979.

[272] Jellinek, H. H. G.; Urwin, J. R., Ultraviolet Absorption Spectra and Dissociation Constants of Picolinic, Isonicotinic Acids and their Amides. *The Journal of Physical Chemistry* **1954,** *58* (7), 548-550.